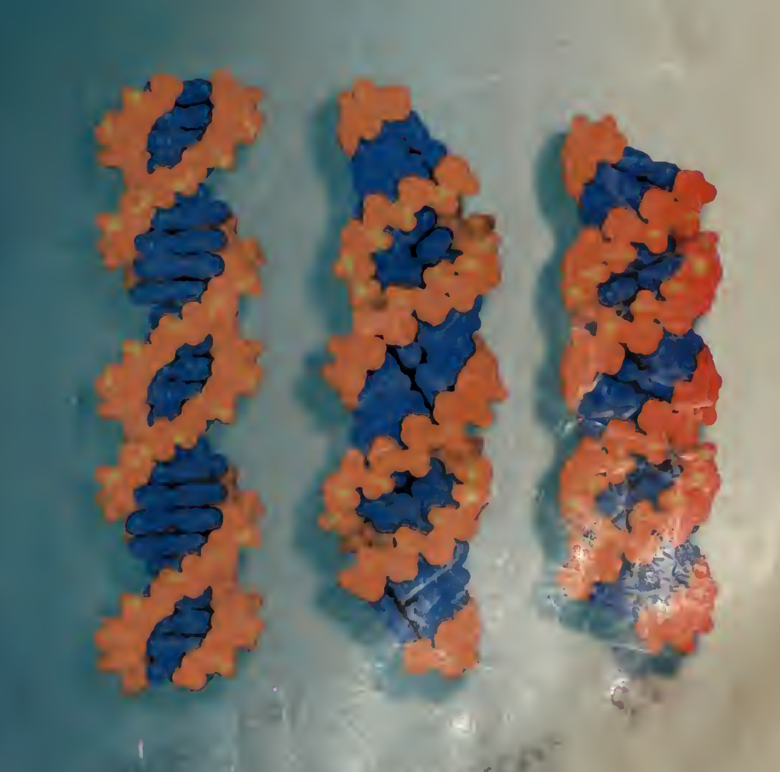
分子生物学

（第二版）

阎隆飞　张玉麟　主编



FEN ZI SHENG WU XUE

中国农业大学出版社

中华农业科教基金会简介

中华农业科教基金会经中国人民银行批准，民政部注册登记，于 1995 年 12 月 20 日成立。基金会得到国家科委、中国人民银行、民政部、农业部等部委的大力支持；得到国内外企业界、知名人士的积极响应。基金会归口农业部管理，接受中国人民银行和民政部监督。

中华农业科教基金会的宗旨是：通过广泛吸收国内外和社会各方面的资金，用以支持中国农业科教事业，补充国家主渠道对农业科技的投入，以加快实施“科教兴农”战略。

中华农业科教基金会的任务是：发展农业科教事业，推动农业科技进步，提高农业劳动者素质，促进中国农业发展和农村经济繁荣。基金会资助农业基础研究、应用研究、试验示范、成果推广和农业科教前沿重大课题的研究；资助有突出贡献和有发展潜力的中青年农业科技人才；资助优秀农业科技著作的出版；奖励在中国农业科教事业中做出重要贡献的个人。

中华农业科教基金会将根据政府制订的农村经济发展规划，定期公布资助方向。资助项目的遴选实行“公开申请，专家评审，民主公正，择优资助”原则。基金会建立严格的筹资、管理和使用制度，公正、合理、规范、科学、有效地使用农业科教基金，向捐赠者公开收支帐目，接受监督。

中华农业科教基金会热忱欢迎国内外企业、社团、各界人士向本基金会捐赠资金，本基金会可根据捐赠者的意愿，设立名人基金、专项基金等。

分子生物学

（第二版）

阎隆飞　张玉麟　主编



中国农业大学出版社



图书在版编目(CIP)数据

分子生物学/阎隆飞,张玉麟主编. —2 版. —北京:中国农业大学出版社,1997.8
ISBN 7-81002-843-X

Ⅰ.分… Ⅱ.①阎…②张… Ⅲ.分子生物学 Ⅳ.Q7

中国版本图书馆 CIP 数据核字(97)第 00994 号

主　　编:阎隆飞　张玉麟
编　　者:李明启　齐顺章　曹淡君　陈祖洁
　　　　　杨礼锐　于嘉林　王荣臣　龚小松
　　　　　阎隆飞　张玉麟

责任编辑:赵玉琴
封面设计:郑　川

出版发行　中国农业大学出版社
经　销　新华书店
印　刷　北京丰华印刷厂印刷
版　次　1997 年 8 月第 2 版
印　次　1997 年 8 月第 1 次印刷
开　本　16　印张 32.25　千字 798
规　格　787×1092
印　数　1～5000
定　价　42.00 元

邮政编码:100094　电话:(010)62892620
社址:北京市海淀区圆明园西路 2 号

再版前言

随着现代科学技术突飞猛进地发展，生物科学也正经历着激烈的变化。在这一转变中，最突出的是分子生物学渗入到各个学科之中，如细胞生物学、发育生物学、遗传学、神经学、生物学、免疫学、病毒学、微生物学、生理学等都几乎无一例外。因此，在高等院校各专业的研究生教学中开设分子生物学课程，作为有关专业的必修课或选修课，对提高研究生的业务水平，是至关重要的。

本书是为了适应当前研究生教学的需要而编写的。它既是一本教材，又是一本专著。第一版已入选为“当代科技重要著作·农业领域”丛书。我们根据第一版在使用时发现一些问题，修订时在章节安排上做了适当地调整，对真核生物基因组结构及其表达调控的内容予以加强。在其余各章节也适当地补充了新内容，使读者能了解分子生物学的新进展。

由于分子生物学的发展极其迅速，文献资料浩如烟海，编者虽尽力吸收各方面的研究材料，但挂一漏万之处在所难免，尚希读者不吝惠予指正。

编　者

1997 年 5 月



目　录



第一章　绪　　论

第一节　分子生物学的含义

生物学经历了一个漫长的研究历程。最早人们从研究动物和植物的形态、解剖和分类开始，以后进一步研究细胞学、遗传学、微生物学、生理学、生物化学，进入细胞水平的研究。到20世纪中叶以来，生物学以生物大分子为研究目标，分子生物学(molecular biology)开始形成了独立的学科，这是对生物界的认识不断深入的过程。

自从1838年Schleiden和Schwan证明动物和植物都是由细胞组成，Virchow(1858)提出细胞学说(the cell theory)之后，细胞学的研究得到迅速发展，随之遗传学原理也得到揭示。同时生理学和生物化学也随之兴起，以细胞为主要材料，进行深入地研究，使生物学的探索进入了细胞水平。

随着物理学和化学渗入生物学的研究领域，人们对细胞的化学组成的了解日益深化，我们对构成细胞的生物大分子，主要是蛋白质及核酸在生命科学中所起的作用有了深刻的认识。Sanger利用纸层析和纸电泳技术于1953年第一次揭示出胰岛素的一级结构，开创了以后数千种蛋白质序列分析的先声。不久Perutz及Kendrew(1953)利用X射线衍射技术解析了肌红蛋白(myoglobin)和血红蛋白(hemoglobin)的三维结构，使人们第一次能够洞察生物大分子的空间结构，从而了解这些蛋白质在运送分子氧中的特殊作用。这些研究结果使人们第一次能从分子水平了解生命物质的结构与功能。生物大分子结构与功能的研究乃成为生命科学中最重要的课题。

在核酸方面随着核酸化学研究的进展，人们揭示出核酸化学的许多规律之后，Watson和Crick(1953)共同提出了脱氧核糖核酸(DNA)的双螺旋模型(double helix model)。这个模型为揭开遗传信息的复制和转录的秘密铺平了道路。随后Crick又提出了中心法则(central dogma)，明确了遗传信息传递的规律。从此以后，核酸的分子生物学得到了异乎寻常的迅速发展，使分子生物学成为生命科学中活力最强的学科。

从广义来讲，蛋白质及核酸等生物大分子结构和功能的研究都属于分子生物学的范畴，也就是从分子水平阐明生命现象和生物学规律。例如，蛋白质的结构、运动和功能，酶的作用机理和动力学，膜蛋白结构功能和跨膜运输等都属于分子生物学的研究内容。

不过目前人们通常采用狭义的概念，将分子生物学的范畴偏重于核酸(或基因)的分子生物学，主要研究基因或DNA的复制、转录、表达和调节控制等过程，当然其中也涉及与这些过程有关的蛋白质和酶的结构与功能的研究。本书也采用狭义的概念对基因分子生物学的基本原理进行讲述。至于专门问题如发育分子生物学、分子免疫学等，则请读者阅读有关专著。

第二节　分子生物学发展简史

随着化学及物理学的渗透，构成生物细胞的生物大分子的结构与功能的研究日益获得突破性的进展。早在 1871 年 Miescher 从死的白细胞核中分离出脱氧核糖核酸(DNA)，迄今已有 120 年的历史。到了 1928 年 Griffith 发现肺炎链球菌(*Pneumococcus*)的无毒菌株与其被杀死的有毒菌株混合，即变成致病菌株。1944 年 Avery 等人发现从致病力强的光滑型(S 型)肺炎链球菌提取的 DNA 能使致病力弱的粗糙型(R 型)转化成 S 型。如果加少量 DNA 酶，这种转化立即消失，但加入各种蛋白水解酶则不能改变这种转化。他们的实验充分证明引起细菌遗传改变的物质为 DNA。

由于核酸化学的研究进展，Chargaff(1949)从不同来源 DNA 测定出 4 种核酸碱基，即胸腺嘧啶(thymidine，T)、胞嘧啶(cytocine，C)、腺嘌呤(adenine，A)和鸟嘌呤(guanine，G)。腺嘌呤与胸腺嘧啶的量和鸟嘌呤与胞嘧啶的量并不相等，即(A＋T)/(G＋C)的比值随不同来源的 DNA 而有所不同。他发现鸟嘌呤的量与胞嘧啶的量总是相等，腺嘌呤与胸腺嘧啶的量相等，即 G═C ，A═T ，这个规律称为 Chargaff 规律。与此同时 Wilkins 及 Franklin(1950～1952)用 X 射线衍射技术测定了 DNA 纤维的结构，它的衍射图像表明 DNA 具有典型的螺旋结构，并且由 2 条以上的多核苷酸链组成。当时，Pauling(1953)曾经提出 DNA 分子具有三股螺旋的设想(见 Nature 1953，171：346～348)。几乎与其同时年青的 Watson 和 Crick 于 1953 年也在 Nature 杂志(171:737～738)上提出了 DNA 双股螺旋模型。DNA 以磷酸糖链形成了双股螺旋，脱氧核糖上的碱基按 Chargaff 规律构成双股磷酸糖链之间的碱基对。这个模型表明 DNA 具有自身互补的结构，根据碱基对原则，DNA 中贮存的遗传信息可以精确地进行复制。他们的理论奠定了分子生物学的基础。

DNA 双螺旋模型已经预示出 DNA 复制的规则。Kornberg 于 1956 年在大肠杆菌(*E. coli*)的无细胞提取液中实现了 DNA 的合成。他从 *E. coli* 中分离出 DNA 聚合酶 I(DNA polymerase I)，能使 4 种 dNTP(即 dATP，dGTP，dCTP 和 dTTP)连接成 DNA。DNA 的复制需要以一个 DNA 作为模板。以后证明 DNA 的复制是一个非常复杂的过程，包含着许多种酶的参与。

DNA 复制在分子生物学中是一个异常重要的问题。Meselson 与 Stahl(1958)用精彩的实验证明，DNA 复制时 DNA 分子的两条链先行分开。他们用 ^{15}N 重同位素及密度梯度超速离心证明 DNA 复制是一种半保留复制。

Crick 于 1954 年提出了遗传信息传递的规律，DNA 是合成 RNA 的模板(template)，RNA 又是合成蛋白质的模板，称之为中心法则：

复制 ⟳ DNA ⇄(转录) RNA —翻译→ 蛋白质

这个中心法则对以后分子生物学的发展起了极其重要的指导作用。

由于多年的研究编码组成蛋白质的氨基酸的遗传密码得到解决，对分子生物学的发展有重要推动作用。蛋白质由 20 种氨基酸组成，而 DNA 仅由 4 种核苷酸构成，按照中心法则，氨基酸与核苷酸的关系如何？Yanofsky 和 Brener(1961)提出了三联体(triplet)的设想，即 3 个

碱基编码一种氨基酸。这个问题经过 Nirenberg 和 Matthai(1963)的努力研究,编码氨基酸的遗传密码终于得到了破译。他们在无细胞系统中加入一定序列的人工合成的多核苷酸,即合成了一定序列的多肽链·充分证明编码 20 种氨基酸的遗传密码。

Khorana(1966)用实验证实了 Nirenberg 提出的遗传密码。Khorana 用有机化学方法合成了多聚脱氧核糖核苷酸,并以它为模板用 DNA 聚合酶 I 合成 DNA 链,然后他以 DNA 为模板用 RNA 聚合酶合成了 RNA 链,二者具有互补的关系。

蛋白质合成是分子生物学的重要课题,它的研究经历了很长的过程。早在 1953 年 Zamecnik 及其同事就开始在无细胞系统中利用放射性同位素标记的氨基酸研究蛋白质合成过程,发现蛋白质合成的场所为核糖体(ribosome)。他们还证明蛋白质合成需要 ATP 作为肽链形成的能源。氨基酸掺入蛋白质之前首先要与转移 RNA(tRNA)结合,它是由氨基酰合成酶(aminoacyl synthetase)催化的。在细胞总 RNA 中 tRNA 约占 10%,RNA 的 85%存在于核糖体(rRNA)中,1960 年以后利用 T4 噬菌体感染 *Escherichia coli* 作为系统,噬菌体侵染细菌后,寄主的 RNA 合成被中止,只有 T4 DNA 被转录成 T4 RNA。令人惊奇的是 T4 RNA 的碱基组成与 T4 DNA 非常相似,但它并不与 rRNA 结合形成核糖体。后来这种 RNA 携带 DNA 的信息转移到核糖体上合成蛋白质,故称为信使 RNA(messenger RNA,mRNA)。mRNA 约占总 RNA 的 4%。继 mRNA 被发现之后,Hurwitz,Stevans 及 Weiss 等人发现了 RNA 聚合酶,这种酶以 DNA 为模板利用 ATP,GTP,CTP,UTP 等合成 RNA,这就是转录(transcription)过程。

在细胞中蛋白质合成是受到控制的。例如,*E. coli* 中的 β-半乳糖苷酶(β-galactosidase)的含量就随着对它的需要而变化。当乳糖存在时它的含量就高,将乳糖分解成葡萄糖和半乳糖,而在乳糖不存在时,细菌合成的 β-半乳糖苷酶则极少。Monod 和 Jacob 于 50 年代末对此问题做了详细研究,提出了操纵子学说(operon theory),指出在操纵子中存在调节基因(regulatory gene),它可以产生阻遏蛋白(repressor),在乳糖不存在时阻遏蛋白就关闭结构基因,使之不能合成半乳糖苷酶。

DNA 是一个长链的生物高分子,在研究 DNA 重组、表达质粒的构造及它的碱基序列分析之前往往需要将 DNA 分子切割成为较短的片段,这就需要一种酶来完成。Smith 于 1970 年从 *E. coli* 中分离出第一个能切割 DNA 的酶。由于它能在 DNA 核苷酸序列的专一性位点上切割 DNA 分子,他将这种酶称为限制性酶(restriction enzyme)。以后很多种限制性酶陆续被分离出来,目前已有数百种限制性内切酶作为商品出售,给分子生物学研究带来极大的方便。

按照中心法则信息传递的方向是从 DNA 到 RNA,再从 RNA 到蛋白质。但是 RNA 在反转录病毒(retrovirus)中并非如此,它们能以 RNA 为模板合成单链的 DNA,然后再以这条 ssDNA(single-strand DNA)为模板合成互补 DNA(complementary DNA,cDNA)。以 RNA 为模板催化 DNA 合成的酶称为反转录酶(reverse transcriptase),它是由 Temin 和 Baltimore(1970)首次分别发现的。现在,我们可以利用反转录酶以分离得到的 mRNA 为模板合成 cDNA,从而进行基因结构及其表达的研究。

限制性内切酶的分离成功使得重组 DNA 成为可能。在此以前已经发现细菌中存在 DNA 连接酶(DNA ligase)。它能将被限制性酶切割的 DNA 片段连接在一起,1972 年 Berg 首次将不同的 DNA 片段连接起来,并且将这个重组的 DNA 分子有效地插入到细菌细胞之中,重组的 DNA 进行繁殖,于是产生了重组 DNA 的克隆(clone)。Berg 是重组 DNA 或基因工程技术

的创始人，于1980年获得了Nobel奖。

在研究分子生物学中了解DNA的核苷酸的排列顺序无疑是非常重要的。Sanger(1977)及Gilbert(1977)分别用与测定蛋白质序列截然不同的方法解决了DNA分子中碱基序列(DNA sequence)的复杂问题。Sanger采用的是酶法，而Glibert采用化学法。他们都将DNA分子中碱基的序列准确地测定出来，使我们对基因甚至基因组的结构得到了解。

自Summer(1936)证明酶是蛋白质以来，已有50多年的历史，人们一直认为酶是蛋白质。但是近年有一个惊人的发现，即一些RNA也具有催化功能。Cech于1986年发现四膜虫(*Tetrahymena*)的核糖体RNA能够自我剪切。mRNA中的内含子能被RNA本身的催化作用准确无误地切除。这种内含子衍生出来的RNA在特定位点催化RNA链的剪切和连接，而其自身并不被消耗，完全符合酶的性质。这种催化剂被定名为核酶(ribozyme)。这一发现使人推测在生物进化的早期可能先形成RNA，然后以RNA为模板形成DNA。DNA后来代替RNA作为遗传物质，它的双螺旋结构比RNA单链更稳定，适宜于遗传物质的存贮，而RNA在核糖体中仍保留着催化性质。

由于分子生物学的研究对生命科学的发展起着巨大的推动作用，受到国际科学界的高度重视，许多位分子生物学家获得了诺贝尔化学奖或生理学奖。现在将历年来分子生物学家获得诺贝尔奖的情况简单介绍于表1中。

分子生物学从开始到如今只有50多年的发展历史，在人类文明史中只是短暂的一段，却使生物学发生了巨大变化，其进展可谓极其迅猛。由于无数分子生物学家的刻苦研究，使我们现在不但能从分子水平上了解DNA的结构、复制、转录和表达的详尽过程，而且对某些重要生物如果蝇(*Drosophila*)或拟南芥(*Arabidopsis*)复杂的发育过程有了深入地了解，使生物科学进入了一个新阶段，这在过去是办不到的。上面介绍的只是分子生物学的少数重要成就，我们不可能作全面详细介绍。关于分子生物学今后的展望，我们将在第四节加以讨论。

表1 分子生物学家获得诺贝尔奖一览表

人 名	年 代	获奖内容
Kornberg, A.	1959	DNA的复制
Watson, J. D.	1962	DNA双螺旋结构
Crick, F. H. C.		
Wilkins, M.		
Monod, J.	1965	操纵子学说
Jacob, F.		
Nirenberg, M.	1968	遗传密码的解析
Khorana, H. G.		
Holly, R.	1968	tRNA的结构
Temin, H.	1975	反转录酶的发现
Baltimore, D.		
Berg, P.	1980	DNA重组，DNA序列分析
Gilbet, W.		
Sanger, F.		
Cech, T.	1989	核酶(ribozyme)的发现

第三节 分子生物学在生命科学中的位置

分子生物学是从生物化学发展出来的一门学科。由于蛋白质及核酸生物化学的研究逐渐深入到这些生物大分子结构和功能的研究中，物理学家利用X射线衍射技术使蛋白质及核酸大分子的结构得到了解析，揭露出它们的三维结构，从而提出了脱氧核糖核酸(DNA)双螺旋模型。此后核酸的分子生物学得到了迅速发展，形成分子生物学一门独立的新学科。但是，分子生物学并不能与生物化学相分离，而是关系日益密切。例如，国际生物化学协会(The International Union of Biochemistry)现已改名为国际生物化学与分子生物学协会(The Interna-

tional Union of Biochemistry and Molecular Biology)，足以表明二者的关系十分密切。此外，近年来蛋白质与DNA结合的研究日益受到重视，成为晶体学(crystallography)的重要内容，这个问题深入探讨将使分子生物学的重要机理，如复制、转录、翻译、调控的本质得以充分阐明，也就是蛋白质和酶究竟是如何使复杂的复制、转录及翻译过程进行的。DNA虽然可以复制，但是DNA本身却不能复制。RNA可以转录，但RNA本身却不能转录。复制和转录都是由蛋白质和酶实现的。翻译过程也是由核酸和蛋白质完成的。由此可见，核酸和蛋白质的关系是相辅相成的。

目前分子生物学的研究虽然仍以DNA重组技术为主要手段，但是愈来愈多的事实表明，蛋白质和酶的研究在分子生物学研究中的重要地位，例如，研究基因的表达问题，必然涉及到基因表达的产物，也就是蛋白质和酶。而要想深刻阐明所表达的蛋白质和酶，就必须彻底了解其结构和功能。因此，在分子生物学的研究中，蛋白质的纯化、一级结构、晶体的三维结构、溶液构象、光谱性质、酶的动力学等就成为必须研究的内容，这样才能对基因表达的产物做出确切的解释。

分子生物学与微生物学的关系是密不可分的。早期分子生物学的研究对象都是原核生物，特别是大肠杆菌(*E. coli*)。大肠杆菌的DNA和RNA及其复制、转录、翻译和调控过程我们已经了解得非常清楚。甚至我们现在用于分子生物学研究的各种质粒(plasmid)和限制性内切酶(restriction endonuclease)大都是*E. coli*中分离出来的。现在研究基因表达也往往是先用*E. coli*作为受体菌。难怪有人认为，目前的分子生物学主要是"*E. coli*的分子生物学"。当然，目前分子生物学的研究对象早已不限于微生物，动物和植物的分子生物学也已取得很大进展。由于微生物学与分子生物学的结合，许多重要的微生物学课题已经达到了分子水平，如生物固氮(biological nitrogen fixation)和植物微生物相互关系(plant-microbe interaction)的机理已日渐明朗。

遗传学无疑是与分子生物学关系异常密切的学科。分子生物学的发展极大地丰富了遗传学的内容，产生了分子遗传学(molecular genetics)。分子遗传学与细胞遗传学、生化遗传学、数量遗传学、群体遗传学等，构成遗传学的重要内容，并且成为遗传学的重要部分。分子生物学极大地丰富了基因学说(gene theory)，不但大量基因的序列分析已测定清楚，而且人类基因组(human genome)的结构也作为重大课题开展了研究。

细胞生物学与分子生物学也有密切的关系。现在，细胞生物学已经不限于从细胞水平研究细胞的形态、结构与功能，而进一步从分子水平探讨构成细胞各种组分的基因及其表达，于是产生了分子细胞生物学(molecular cell biology)的新学科。这个学科正在蓬勃发展，将会改变细胞生物学的面貌。

发育生物学研究动植物的生长发育，是生命科学的重要内容。过去的组织学和胚胎学偏重于动植物结构和发育的形态描述。实验胚胎学虽然对生物的发育进行了细胞水平的研究，但对发育的实质难以揭示。近年来分子生物学渗入到发育生物学以后，发育生物学的面貌大为改观。现在我们已经认识到不论低等真核生物或高等生物，它们的发育都是受基因组中DNA控制的，并且按照特定的时间和空间顺序(temperal and spacial sequence)依次表达出来，从而认识生物发育过程的实质。

目前生理学的研究也与分子生物学密切结合起来，不论动物生理学或植物生理学都是如此。例如，神经系统分子生物学的研究造成神经生物学的革命性变化。乙酰胆碱受体的分离纯

化和基因克隆已经完成。神经通道也已得到纯化和基因克隆。分子生物学渗入神经科学已产生了分子神经生物学。在植物生理学方面光合作用的研究已进入分子水平。一些植物的叶绿体DNA 的全序列测定已经完成。光合作用有关的基因如光系统Ⅰ,Ⅱ,电子传递链,二磷酸核酮糖羧化酶/加氧酶(Rubisco)等基因的表达调控大量的研究都已完成了,使光合作用进入了一个新阶段。

不仅上述学科与分子生物学有密切关系,生物科学的许多学科也都进入分子生物学的研究领域,如分子免疫学、分子药理学、分子病理学、分子分类学等等,都从分子水平进行探索,使这些学科进入更高的阶段。

第四节　分子生物学的现状和展望

分子生物学的现状和发展前景是非常令人鼓舞的。90 年代以来分子生物学在理论和技术方面都取得了重要进展,在 DNA 的复制、修复、转录、翻译和调控的分子机理等方面都得到了进一步的阐明,如拓扑异构酶 I(topoisomerase I)的晶体结构、核糖体(ribosome)结构的研究使得我们对 DNA 的复制、翻译等的认识比过去更深入了。

分子生物学正在渗入到生物科学的有关学科之中,使得这些学科的面目大为改观。首先,在发育生物学方面,生活周期短的一些动植物如线虫(*Caenorhabditis elegans*)、果蝇(*Drosophila melanogaster*)、拟南芥(*Arabidopsis thaliana*)已成为发育生物学的重点研究对象,它们的发育过程很多已从分子水平得到了解。例如,果蝇从卵发育成成虫的遗传信息是编写在基因组中的。一切性状都是由相关的基因所决定。在发育中基因的表达在时间和空间顺序上受到控制,如果蝇不能在幼虫时表达腿部的基因,只能在蛹期才能长出腿,也不能在蛹期产生卵,只能在雌性的成虫产生卵细胞。植物也是如此,它的生长发育过程也是在时间和空间上依靠各种基因依次表达的。这样,生物体的各种代谢过程才得以有条不紊地进行。例如,植物开花的基因十分复杂,花的各种器官由许多基因控制着,花瓣的颜色由多种次生物质代谢产物(如黄酮类色素 flavanoid pigments)所决定,它要在开花期才表达形成,而不会在幼苗期表达。现在黄酮类色素生物合成的途径及有关基因已基本研究清楚。

分子生物学与细胞生物学关系密切,已形成一门新的分子细胞生物学(molecular cell biology)。许多细胞生物学问题如细胞分裂、细胞骨架(cytoskeleton)、细胞因子(cytokine)的研究都进入了分子水平。

免疫学与分子生物学结合,产生了分子免疫学(molecular immunology)。病理学与分子生物学结合,产生了分子病理学(molecular pathology),其中病毒学与分子生物学结合,就是分子病毒学(molecular virology)。目前,分子生物学几乎已渗入到所有生物科学的各个领域,甚至最古老的动物和植物的分类学也开始采用分子生物学研究物种的亲缘关系,于是有分子系统学(molecular systematics)的出现。

自 1990 年以来在生物科学中兴起了一门新的结构生物学(structural biology),它专门研究生物大分子的空间结构和功能。由于结构生物学在发展生物科学的重要性,著名的 Nature 杂志从 1994 年专门出版了 Nature Structural Biology 月刊,充分登载结构生物学的研究成果。Current Biology 公司则出版了 Current Opinion of Structural Biology,Structure 及 Macromolecules 等刊物,分别发表有关蛋白质及核酸的三维结构。目前,生物大分子三维结构的研究

进展极快，在全世界范围业已达到平均每天能解析出三种蛋白质晶体结构的速度，而 30 年前，10 年间只解析出一种蛋白质结构，结构生物学这样的高速度发展对生物科学做出了重大贡献。高分辨的蛋白质晶体结构使我们清晰地看到蛋白质分子中多肽链如何折叠成 α 螺旋、β 折叠片、γ 转角以及整个分子的折叠情况，甚至蛋白质（酶）与配基（ligend）的结合情况。在结构生物学中 DNA 与蛋白质的相互作用是一个重要领域，它对分子生物学的理论研究至关重要。例如，在核酸结构中除去双股 DNA（DNA duplix）外，还发现了三股 DNA（DNA triplex）和四股 DNA（DNA tetraplex），三股 DNA 在抑制真核基因表达中有重要作用，而四股 DNA 存在于染色体的端粒（telomere）中，有稳定染色体结构的作用。目前，晶体学家已经解析出大量与 DNA 相互作用的蛋白质，如各种限制性内切酶（*Eco*RI，*Bam*HI），各种阻遏蛋白（repressors，CAP，trp，Cro，434，Arc，MetJ）、DNA 修复蛋白（DNA repair protein）、TATA box 蛋白、组蛋白（histone）、转座酶（transponase）等。拓扑异构酶 I（topoisomerase I）的三维结构也已解析出来，它是一个很特殊的酶，具有环状结构，它能在瞬间打断一股 DNA，将另一股 DNA 穿过断裂处，然后又封闭裂口，使两股 DNA 重新组合。拓扑异构酶 I 在 DNA 复制、转录及重组中起重要作用。此外，分子生物学对结构生物学的发展也有很大帮助。从动植物、微生物细胞中分离纯化蛋白质是一项十分艰苦的工作，由于大多数蛋白质在细胞中的含量都很低，要想分离纯化出某种蛋白质，进行晶体培养，是相当困难的。现在，我们可以将编码某种蛋白质的基因用分子生物学技术进行克隆，用聚合酶链式反应（PCR）扩增，然后与融合蛋白质粒（fusion protein plasmid）构建成表达质粒，用它转化大肠杆菌（*E. coli*）。然后进行发酵培养，即可得到大量含有该种蛋白质的菌体。从这些菌体中即可分离纯化出足够数量的该种蛋白质，从而培养晶体就容易得多了。

蛋白质工程（protein engineer）技术与分子生物学结合是促进分子生物学发展的一条途径，采用定点突变（site-directed mutagenesis）方法使基因结构发生改变，从而可以改变基因表达产物中的氨基酸残基，就有可能使我们了解蛋白质中每个氨基酸甚至每个化学基团所起的作用。

目前，动植物基因组（genome）的研究已成为分子生物学的重大课题。最近线虫（*Caenorhabditis elegans*）基因组即将测定完成，它由 1 000 万个核苷酸组成。我国科学家正在进行水稻（*Oryza sativa*）基因组的 DNA 序列测定，这项工作的完成必将对植物分子生物学及农业科学做出重大贡献。目前，人类基因组（Human genome）的研究是国际上的重大科学课题，世界上许多国家投入大批人力、物力研究人类染色体 DNA 的全序列。这项宏伟工程的完成将会对分子生物学做出巨大贡献，并将对医学特别是遗传病的诊断治疗做出贡献。

今后，分子生物学在实际方面的发展也是令人乐观的。在医学方面，基因治疗（gene therapy）目前已在研究之中。在动植物方面转基因动植物的培育，也已进入重点研究之列，成为生物技术（biotechnology）的主要内容。将苏云金杆菌毒蛋白（*Bt*）基因转入棉花已获得抗棉铃虫的抗虫棉，并已取得经济效益。利用反义 RNA（antisense RNA）技术延长果实、蔬菜的保鲜期和改变花卉的花色也已取得成果，这些都是分子生物学在实际方面的贡献。随着分子生物学和生物技术的发展，它们在医学和农业方面必将对人类的健康和生活做出更大的贡献。

（阎隆飞）

第二章　生命的物质基础

对生命现象本质的认识，历来存在两种截然不同的看法：生命是一种超自然现象和生命是物质的一种属性。随着物理学和化学对生物学的渗透和融合，作为生物学语言的生物化学出现了。而蛋白质和核酸等生物大分子结构和功能关系的阐明进一步使人们得以从分子水平上来认识生命现象的总体。事实表明，生物机体中物质的转化以及伴随的能量变化，各种化学组分在时空上的相互关系，以及驾驭这一切的遗传信息的传递无一不遵循物理学和化学的基本法则。这些都证明生命不是超越自然的，它的本质即是物质运动的一种状态。

第一节　生命是物质进化的产物

洪荒时代的地球是一个无生命的世界，它笼罩在一种还原性大气中，当时大气的成分主要是 H_2，CH_4，NH_3，N_2，CO，CO_2，H_2S 等。以后随着水分由地球内部外溢，海洋形成了。从此，地球上的化学反应活跃了起来，地球上的物质开始了化学进化。无机化合物藉地热、放电、紫外线、宇宙射线等的能量合成了简单的有机化合物，随后氨基酸、糖、脂肪酸、核苷酸、卟啉化合物等相继形成。地球上出现了非生物合成的有机化合物。随着有机化合物的蓄积，原始生命终于在某种特定环境中出现了，这大约发生在 30 亿年前，即前寒武纪年代，物质进化从此步入生物进化（图 2-1）。最原始的生物是一些异养厌氧生物，它们以环境中的非生物合成有机成分为养料，

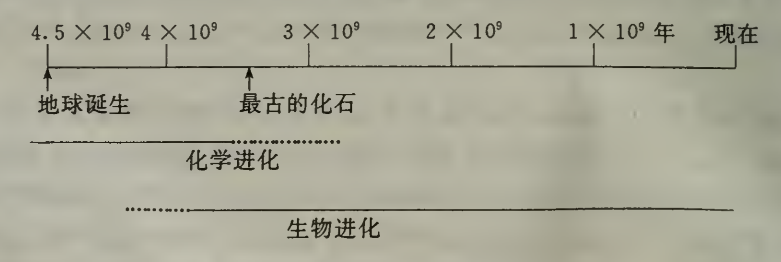


图 2-1　地球上的物质演化进程

通过酵解方式获取生物可以利用的能量形式——ATP。但环境中的有机成分有限，这就限制了原始生物的发展。随着时间的推移，一种光合自养生物出现了，它使还原性大气中出现了氧。氧虽不是生命所必需的，却是高水平的能量代谢所不可少的。氧对生物进化孕育着巨大的潜力。有氧生活虽更能满足高等生物所需的能量，不过氧对原始厌氧生物却具有毒性。大气中氧的出现推动了生物进化，原始厌氧生物终于获得了对氧毒性的防护机制。糖的酵解发展为有氧分解，原核生物发展为真核生物。这些变化为生物进化在遗传上和代谢上准备了条件，生物进化加速了。

生命，不管是其低级或高级形态都显示出与无生命物体明显不同的特点。生命的出现是物质进化的一个转折点。

生命所具有的特征是：

1. 生物体是由许多称为生物分子(biomolecule)的有机化合物组成的。活机体中,生物分子是有组织的,它们按一定的层次形成一高度有序的系统。活机体能维持这一有序性直至死亡。无生命物质则相反,它们总是自发地向无序方向发展。

2. 活机体不断地与环境发生物质和能量的交换。活机体中存在着一个具一定顺序、相互协调、可自我调节的代谢网络,其中各个代谢反应都由相应的酶催化。细胞和机体与环境保持在一个远离平衡态的稳态(steady state)中。与此相反,无生命物质总是趋向于与环境达成平衡。

3. 生物体能够精确地自我复制,即机体的形状、结构和功能等能够代代相传,能够生长、繁殖。与此相反,无生命物质则不能自我生长和繁殖。

什么是生命是一个难以简单界定的问题,它的答案还得从活机体的自身去寻找。

第二节　生命的分子逻辑

活细胞或机体是由许多无生命的有机化合物组成的。但细胞不是一般意义上的口袋,会包容所有相遇的物质,而是有选择的摄取。在物质进化过程中,只是特定生物分子的集聚才导致生命的出现。在物质化学进化的末期,核苷酸、氨基酸、脂肪酸、卟啉等小分子化合物出现了。根据现有资料,人们认为这时可能出现了关键性的一步,即是核苷酸缩聚成了RNA。它是第一个具催化功能的基团,它能催化自我复制。RNA 自我复制中产生的变异体引起了氨基酸的缩聚,蛋白质出现了。以后,DNA 替代 RNA 成为遗传信息的载体,脂类物质形成的半通透性膜使前述的分子聚集体形成为细胞。

上述生命发生过程虽在一定程度上仍是一种假设,犹待进一步验证。不过,这一假设已指明,在细胞和机体中,生物分子是高度组织化的,形成了严密的系统,生命现象是它的总体表现。

一、机体化学组成上的统一性和多样性

活机体,不论是原核或是真核生物都是由核酸、蛋白质、脂质和糖类等生物大分子和一些小分子化合物及无机盐组成的。表 2-1 列出的是原核生物大肠杆菌的分子组成。事实上,所有生物的化学组成都可以分为这几类,这就是其统一性之所在。不同生物在化学组成上的差异不在于分子的类别,而在于每一类分子的质与量。如以大肠杆菌与最简单的真核生物酿酒酵母相比,前者含 1 个染色体,其 DNA 含 4.7×10^6 bp(base pair,bp),而后者含 16 个染色体,其 DNA 含1.4×10^7 bp。不仅如此,DNA 的核苷酸序列也不同。DNA 质和量上的不同决定了细胞所含的 RNA 的不同,主要是mRNA的不同,从而也使所含的包括酶在内的蛋白质的质和量不同,转而影响机体的代谢和其它的化学组成。机体这种代谢上的差异就反映为生物性状上的多样性。生物的多样性是以其统一性为基础的。

表 2-1　大肠杆菌细胞的分子组成

	占总重的百分数	分子的种数
水	70	1
蛋白质	15	3 000
核酸		
DNA	1	1
RNA	6	>3 000
多糖	3	5
脂质	2	20
生物大分子的结构单位和中间体	2	500
无机离子	1	20

二、生物大分子的高聚物特性

核酸(DNA 和 RNA)、蛋白质、多糖和脂质是组成生物体的 4 类生物大分子。DNA 是生物体中遗传信息的原初载体。DNA 通过复制使遗传信息由亲代流向子代,通过转录使特定基因的遗传信息转换成相应的指令——mRNA,后者指导氨基酸按一定的顺序连接成特定的多肽,然后折叠成相应的蛋白质。蛋白质是遗传信息的体现者。核酸和蛋白质合成代表生命活动中遗传信息流动的主线,它驾驭着生命活动的进行。核酸和蛋白质的高聚物特性正是实现这种信息流动的基础,核酸分子的骨架是由核苷酸通过 3′,5′-磷酸二酯键连接成的多核苷酸链,核苷酸是其单体。构成 DNA 和 RNA 的分别是 4 种脱氧核糖核苷酸和核糖核苷酸。不同的核糖核苷酸(和脱氧核糖核苷酸)的区别在于其碱基的差异。蛋白质分子的骨架是由 20 种氨基酸通过肽键连接成的多肽链。20 种氨基酸的区别在于其侧链(R 基)的差异(图 2-2)。这就极大地简化了遗传信息的转化,使其成为 4 种核苷酸和 20 种 α-氨基酸连接顺序间的转换,亦即核酸语言转换成了蛋白质语言。在转录中,DNA 的碱基顺序决定了新合成的 mRNA 的碱基顺序,这是遗传指令的发送。在翻译中,mRNA 上的碱基顺序规定了新合成的多肽链的氨基酸顺序,而氨基酸侧链的结构和性质则决定了多肽链可折叠成的稳定构象和形成相应的功能。这是指令转换为功能的过程。

图 2-2　核苷酸和 α-氨基酸

除了核酸和蛋白质外,多糖和脂质是组成细胞的另两类重要的物质。多糖的单体是单糖。根据单糖的种类,多可分为同多糖(homopolysaccharide)和杂多糖(heteropolysaccharide)。同多糖由一种单糖组成,杂多糖由一种以上的单糖组成。淀粉、糖原和纤维素属同多糖,都由 *D*-葡萄糖组成,但它们的分子内糖苷键的类型有所不同。糖蛋白和糖脂的糖链属杂多糖,它们不仅组成单糖的种类多,并且单糖残基间糖苷键的类型也多样化。这些寡糖链是机体中分子的识别标志。寡糖链的组成和结构也是受遗传信息控制的,它们是在特定酶的催化下形成的。

脂质是一个松散的概念。它包括许多化学上不同的物质,如油脂、磷脂、类固醇、类胡萝卜素和脂溶性维生素等。它们的共同特点是难溶于水。脂质从结构上考虑不属于高聚物,但脂质中的磷脂等是两亲分子(amphipathic molecule),即分子的一部分亲水另一部分亲脂,在水中能藉疏水交互作用聚集成双分子层,是膜的基本组成物质之一。磷脂分子中脂肪酸链的不同会影响它们在膜中排列的规整性,从而影响膜的一些特性,如流动性等。这类似于单体对高聚物的影响。

从以上的讨论,我们不难看出表 2-1 所列的 6 000～7 000 种分子是由 4 种核糖核苷酸,4 种脱氧核糖核苷酸,20 种 α-氨基酸以及少数几种单糖和脂肪酸等组成的,这些生物大分子又进而形成超分子复合物和细胞(图 2-3)。

细胞
超分子复合物　　染色体，核糖体，膜，微管等
生物大分子　　DNA，RNA，蛋白质，多糖等
单体　　核苷酸，氨基酸，单糖，脂肪酸等

图 2-3　细胞的结构层次

三、生物大分子单体在代谢中的地位

各种单体不仅是生物大分子的结构单位，又是合成机体的许多重要组分的前体（图 2-4）。

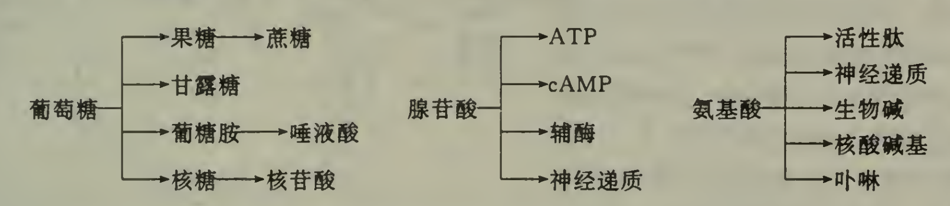


图 2-4　生物大分子单体是一些重要生物分子的前体

生物大分子的单体，尤其是单糖、脂肪酸以及氨基酸也是机体中间代谢的重要环节，它们循不同而又相互连接的途径逐步降解并汇集到少数共同产物，如 α-酮戊二酸（C_5），草酰乙酸（C_4），丙酮酸（C_3）和乙酰辅酶 A（C_2），然后进入三羧酸循环并通过呼吸链生成二氧化碳（C_1）和水（图 2-5）。这是一放能过程，所释放出的能转化为代谢中的能量通货——ATP。分解代谢和合成代谢是新陈代谢的两个基本方面。从总体说，分解代谢是放能过程，合成代谢是需能过程。分解代谢释放出的自由能使 ADP 和无机磷酸（Pi）形成 ATP，合成代谢所需的能可由 ATP 裂解为 ADP 和 Pi 来提供。ATP 是这两大过程间能量转移的基本环节（图 2-6）。

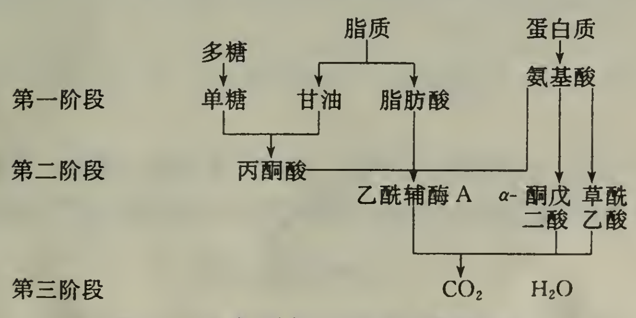


图 2-5　多糖、蛋白质和脂质的降解过程

图 2-6　合成代谢和分解代谢间的能量转移

四、活机体是一远离平衡态的开放系统

分解代谢和合成代谢间的能量转移并不说明两者间存在平衡，而只是说机体中ATP处在稳态中。在活机体中，ATP转化为ADP和Pi的速率导致ADP和Pi重新合成ATP，故ATP的总浓度趋向于不变，但细胞和机体是远离平衡态的。

细胞的各部分是不均匀的，各类生物分子的分布存在着差异，各部分间也存在着分子和离子的浓度梯度以及电位差等，正是这种差别使细胞各部分间存在物质和能量的流动。同样，细胞、机体与环境间也存在着差异，这种差异较细胞内各部分间的差异更为巨大。一膜之隔，外部是无生命的，死气沉沉；而内部是有生命的，生机盎然。内外一旦平衡，死亡立即降临。

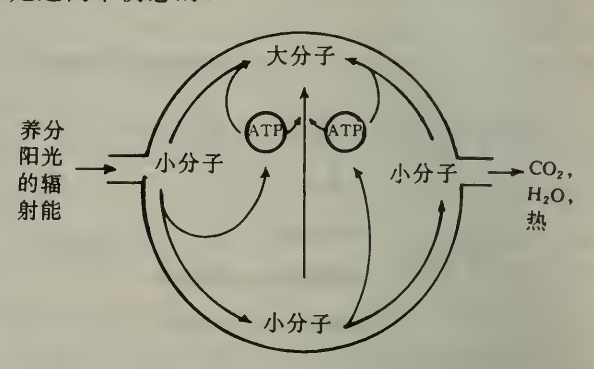


图2-7　细胞为一开放系统

已如上述，活机体是一高度有序的系统，这是生命的基础所在。机体为了保持自身的这种有序性就必须从环境吸收淀粉、脂肪和蛋白质等有机养分，再通过分解代谢取得自由能，也可从阳光摄取辐射能。在此同时，机体向环境排出CO_2，H_2O等简单分子以及在等温、等压条件（这是机体的工作环境）下难于利用的热。细胞为一开放系统（图2-7）。

细胞的代谢构成了一个复杂的网络，各个代谢反应都是酶促反应。代谢反应基本上都是可逆的，但在活细胞中，由于反应产生的产物很快被移去，故反应难以达成平衡。这使细胞中的代谢物保持一定流向。远离平衡态是活机体的一个基本特征，细胞是一个远离平衡态的开放系统。

第三节　生 物 分 子

生命是物质进化到达一定阶段后的产物。在此之前存在于地球上的矿物质并不能导致生命的出现，只有当核苷酸、氨基酸、脂肪酸、单糖等形成后，生命的诞生才具备了条件。

一、生命的化学元素

生命的产生具有明显的化学背景。在多于90种的天然存在元素中，只有约30种是机体必需的，其中碳、氢、氧、氮占多数细胞质量的99%以上。假如以地壳、海水和人体的元素组成做成一比较（表2-2），我们不难发现人体的元素组成相对更接近于海水。这和构成细胞的有机化合物形成于陆地，而生命则发生于海水的假设相一致。

水是机体中一种大量存在的化合物，它占机体重量的70%或更多。许多证据表明，生命发生于水中。水影响着细胞中生物分子间相互作用，水的电离特性也影响着生物分子的功能，水也使细胞质呈现出一定的结构性。水对生命现象具有很大的重要性。除了氢、氧外，碳构成了生物分子的骨架。碳具有一个非常独特的性质，即碳原子能连接成各种稳定的链或环，所以，生物机体不选择与碳同族的、在地壳中含量丰富的硅。生物分子除碳以外还含氢、氧、氮、磷、硫，

这 6 种元素有一个共同的特点，即都是轻元素，都能形成稳定共价键，除氢外还都能形成重键（双键和三键）。它们形成了具不同结构和性能、适合表现出生命现象的化合物。

表 2-2　地壳、海水和人体组成元素的丰度

	海水(%)		人体(%)		地壳(%)
H	66	H	63	O	47
O	33	O	25.5	Si	28
Cl	0.33	C	9.5	Al	7.9
Na	0.28	N	1.4	Fe	4.5
Mg	0.033	Ca	0.31	Ca	3.5
S	0.017	P	0.22	Na	2.5
Ca	0.0062	Cl	0.08	K	2.5
K	0.0060	K	0.06	Mg	2.2
C	0.0014	S	0.05	H	0.22
		Na	0.05	C	0.19
其它	<0.1	其它	<0.01	其它	<0.1

除了上述 6 种元素外，细胞还含有许多微量元素：铁、铜、锰、锌、钴、钼、硒、钒、镍、碘、镁等。

这些元素的含量虽甚微，但它们对生命活动却具有十分重要的功能。其中许多是酶的不可缺少的组分，有的是氧化还原中的电子载体，有的参与代谢的调节，有的是呼吸作用和光合作用所不可缺少的，等等。

二、生物分子是碳的化合物

生物分子是碳的化合物，碳构成了它们的骨架，在生物分子中，碳原子可通过单键相连接，也可以通过双键（生物分子很少含三键）相连接；碳原子可以连接成链，也可以连接成环。分子中，饱和碳原子的 4 个价键分布于空间，故呈立体型；以双键连接的碳原子上的 4 个价键分布于同一平面中，故呈平面型（图 2-8）。这是生物分子的结构呈现出多样性的基础。

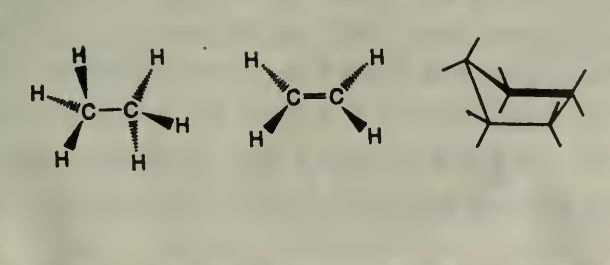




图 2-8　有机化合物的分子模型

碳氢化合物是有机化合物的母体，化学上稳定，但当它们的分子中引入一定的官能团后，化学活动性就相应增加了。生物分子中常见的官能团有羟基、羰基、羧基、氨基、巯基等（图 2-9）。

$R^1—CH_2—O—H$

$R^1—CH_2—C(=O)—H$

$R^1—CH_2—C(=O)—CH_2—R^2$

$R^1—CH_2—C(=O)—OH$

$R^1—CH_2—C(=O)—O—CH_2—R^2$

$R^1—CH_2—O—CH_2—R^2$

$R^1—CH_2—NH_2$

$R^1—CH_2—C(=O)—NH_2$

$R^1—CH_2—NH—C(=NH)—NH_2$

$R^1—CH_2—C{=}CH—N{=}CH—NH$（咪唑环）

$R^1—CH_2—S—H$

$R^1—CH_2—S—S—CH_2—R^2$

$R^1—CH_2—O—P(=O)(OH)—OH$

图 2-9　一些常见的官能团

生物分子除碳环外还常含杂环(由碳及非碳原子构成的环状结构)，它们在生命活动中起着重要作用，如核酸的嘧啶和嘌呤碱基，氨基酸侧链中的咪唑和吲哚基以及血红素和叶绿素中的卟啉等。杂环化合物常含共扼双键，故呈扁平型。

生物分子的各种骨架以及它们具有的不同官能团使它们在结构和功能上显示出高度的多样性，从而满足生命活动的需要。

三、生物分子的构象和构型

关于一个化合物的分子结构，我们不仅要知道其构成原子的种类和数目，相互间的连接方式和连接顺序，还要确定原子在空间的排布。分子中组成原子的空间排布对生物分子，尤其是生物大分子十分重要，因为三维的细胞是由三维的生物大分子组成的，而生物分子间高度专一的相互作用也是通过空间关系实现的。构象和构型就是指分子中组成原子在空间的相互关系。

1. 构象　乙烷分子含 2 个碳和 6 个氢原子，这 6 个氢原子在空间的关系不是固定的，这是由于两个碳原子可以单键为轴心而相对转动。不过两个碳原子的相对转动不是完全自由的，这是由于当碳原子作相对转动时，非键合的氢原子间会发生不同程度的干扰，从而使分子的势能发生改变(图 2-10)。图中表示的只是乙烷分子中氢原子空间排布的两种极端方式，一种称为重叠式；另一种称为交叉式。重叠式的势能最高，交叉式最低，其它方式则介于两者之间。如乙烷那样，由于有关原子相对转动所造成的相应原子或基团在空间的不同排布称为构象(conformation)。

乙烷分子只含两个碳原子，它的构象变化较为简单。但当乙烷分子的两个碳原子各有一个氢为甲基取代时(即形成丁烷)，情况就复杂得多。如仅取第二和第三两个碳原子，它就可以形成如图 2-11 的几种主要构象，如再考虑第一和第二及第三和第四碳原子的相对转动，则不仅构象变化更为复杂，并且各个极端构象间的势能差也将扩大。不同构象能量上的差异来自有关原子或基团的大小和性质，不同大小的原子或基团会产生不同的空间障碍，而基团的性质如是否具有极性会产生排斥或吸引。表 2-3 列出了一些原子的大小，其中共价半径代表成键时该原子的半径，而范德华半径则代表从一个原子的中心到它所接触的非键合原子的距离。共价半径之和即为该两原子间的键长，而范德华尔半径则决定分子可达到的紧密程度。

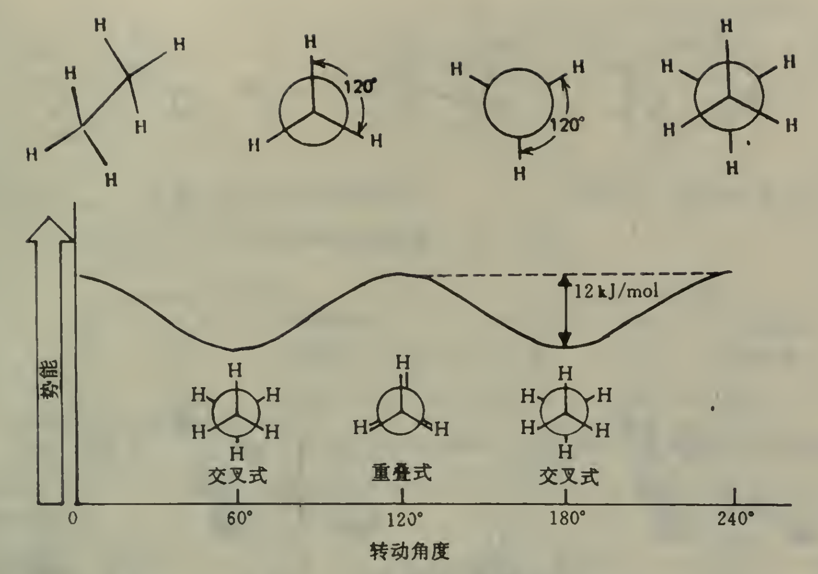


图 2-10　乙烷构象体的能量变化

上图：乙烷构象的 Newman 投影式；下图：乙烷的构象及其能量变化。

除了上述的含单键的分子可以形成各种构象外，一些具环形结构的分子也可采取不同的构象，如环戊烷、环己烷等。己糖常以环形结构的椅型存在。β-D-葡萄糖还会由于羟甲基这一较大基团在空间可采取的取向而形成两种构象，它们可以相互转化，但更多的是以势能较低的一种存在（图 2-12）。

表 2-3　一些元素的范德华半径和共价半径单键

元素	范德华半径（nm）	共价半径（nm）
H	0.10	0.030
O	0.14	0.074
F	0.14	0.071
N	0.15	0.073
C	0.17	0.077
S	0.18	0.103
Cl	0.18	0.099
P	0.19	0.110
Br	0.20	0.114
I	0.22	0.133

2. 构型　当一个碳原子上连接 4 个不同的原子或基团时，这 4 个原子或基团在空间可能有两种排布顺序，即一种按右手性方向排列；另一种按左手性方向排列。含有这一结构特点的两种分子互为镜像，或如左右手关系一样，不管在空间如何转动它们都不能重叠，故它们称为不对称分子或手性分子（图 2-13）。连有 4 个不同基团的原子称为不对称中心或手性中心。由于手性的存在而造成的基团不同空间排布称为构型（configuration）。α-氨基酸（甘氨酸除外）可以分为 *L* 型和 *D* 型就是一例。例如，*L*-丙氨酸和 *D*-丙氨酸的分子式、分子量相同，所含的基团也相同，故而化学性质相同；但α-碳上4个基团

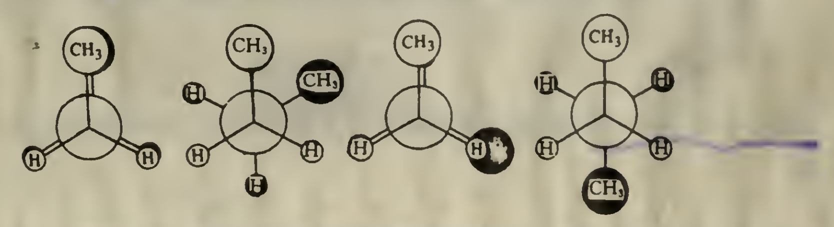


图 2-11　丁烷的几种主要构象

羟甲基，羟基位于平伏键上　　　　　　羟甲基，羟基位于直立键上

图 2-12　β-D-葡萄糖的两种异构体

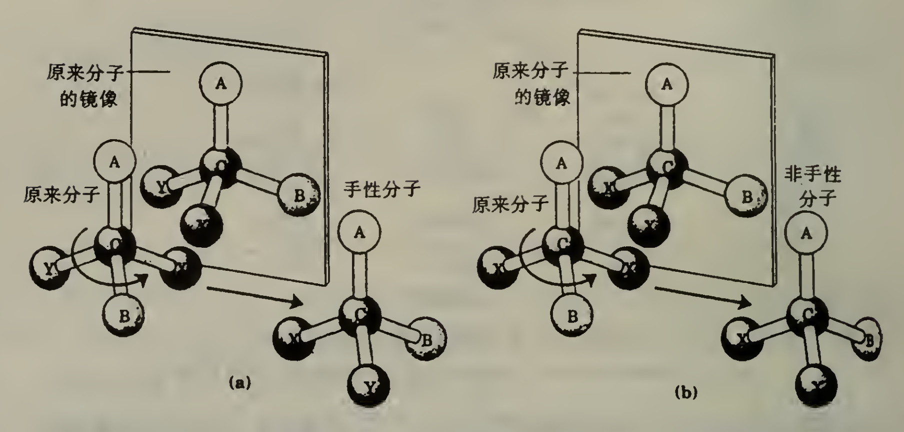


图 2-13　手性和非手性分子

(a)为手性分子；(b)为非手性分子。

的空间排布不同，结果造成它们旋光性的不同。除 α-氨基酸外，乳酸、甘油醛、糖等都是不对称的。

除了不对称原子外，分子中的双键也可造成有关基团在空间的不同排布。丁烯二酸的 C-2，C-3 上连接着氢和羧基，由于 C-2 和 C-3 以双键相连从而不能作相对转动，故而氢和羧基在空间可以有二种排布方式，即两种构型，结果形成马来酸和延胡索酸。

```
H         H              HOOC        H
  \     /                    \     /
   C═C                        C═C
  /     \                    /     \
HOOC     COOH              H        COOH
  马来酸                     延胡索酸
```

构象和构型虽然都涉及基团的空间排布方式，但不同构象可通过绕单键旋转而相互转换，而构型则不能转换除非发生相应共价键的断裂。

构象和构型对理解生物分子的结构和功能具有重要意义。

第四节　生物分子的强键和弱键

一、强键和弱键

生物分子是共价化合物，共价键是通过成键原子间电子对的共享形成的，这是一放能过程。如：

$$H^{\cdot} + {}^{\cdot}H \longrightarrow H—H + 热$$

相反，为使氢分子裂解为氢原子就需加入同样数量的能。这一数量的能一般称为键能，可用作

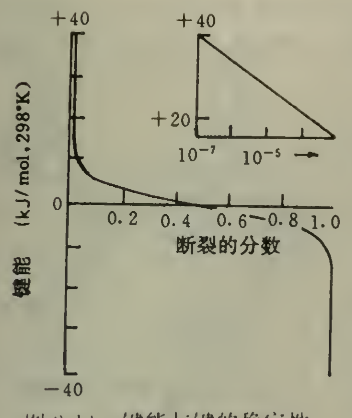


图 2-14　键能与键的稳定性


表 2-4　生物分子中常见共价键的强度

键的类型	键能(kJ/mol)	键的类型	键能(kJ/mol)
O—H	461	C—N	293
H—H	435	C—S	260
P—O	419	N—O	222
C—H	414	S—S	214
N—H	389	C═O	712
C—O	352	C═N	615
C—C	348	C═C	611
S—H	339	P═O	502

键强弱的量度。但对含一个以上相同键的分子而言，情况就有不同。如甲烷分子含 4 个 C—H 键，这 4 个 C—H 键如逐次裂解，每一 C—H 键裂解所需的能量并不完全相同，它们称为裂解能。这时键的强度要用裂解能来表示。不过在多数情况下，我们用它们的平均值来衡量键的强度。

共价键的键能一般都在 200 kJ/mol 以上(表 2-4)，强度都较大。尤其是组成生物分子的都是轻元素，形成的键都很稳定。但生物分子中(间)常存在其它的作用力，如氢键、盐键、范德华力和疏水交互作用，它们的强度常小于 20 kJ/mol，故称为弱键。

从原则上言，强键也不是无限稳定的，在某一瞬间总还会有一定数量的键会断裂，如断裂后的原子并未脱离原来的位置，键随后又可重新形成，键的形成和断裂处于动态平衡中。平衡时，已断裂键所占的分数随键能的大小而有不同(图 2-14)。如某一键的键能为 6 kJ/mol，室温时，已断裂键的分数约为 0.1；如键能达到 200 kJ/mol，这一分数值仅约为 10^{-35}，即原子实际上都以键合状态存在。弱键的一个特点是：在各个瞬间总有一定数量的键发生断裂。弱键对生物分子的结构和功能具有重要意义，下面就分别加以讨论。

二、氢　键

当一个共价结合于强电负性原子上的氢与另一电负性原子接近时，两电负性原子就可能共享一个质子而形成氢键(hydrogen bond)。水分子间的氢键就是一例。生物体中最重要的氢键存在于氧氮、氧氧和氮氮之间(图 2-15)。

图 2-15　一些重要的氢键

在氢键中，质子位于两个电负性原子间从而屏蔽了负电荷的交互作用。如构成氢键的三个原子不处在同一直线上，屏蔽作用就会减弱，甚或氢键就不能形成。

氢键的键能较小，约为 20 kJ/mol。个别的氢键虽较弱，但生物大分子内氢键的数量往往很多，故总体言，其强度仍是可观的。

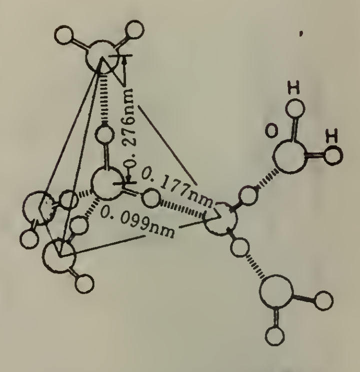


图 2-16　水分子间的氢键

细胞为一含水系统。水对细胞的结构和发挥功能具有重要意义。水是一个高度缔合化合物，即分子间存在着氢键。氢键使水，即使在液态时也具有一定的结构性(图 2-16)。氢键使水具有强内聚力，高气化热和熔化热。它为细胞提供了一个稳定的内环境。水分子不仅相互间形成氢键，也与许多生物分子形成氢键，从而影响生物分子的结构和功能。水的强极性使它成为一种对极性化合物和离子型化合物的良好溶剂，而非极性化合物都难溶于其中。这是形成膜结构的基础。水也是一种电离性溶剂，它影响着溶质分子的离解，水在细胞中不仅是一种介质，也是一种反应物和产物，它参与了许多代谢反应。

三、离　子　键

离子键代表着正负离子间的引力。在真空或空气中，它们属强键，但在水溶液中，它们的强度大为减弱而属于弱键。这是由于水分子为一偶极，水分子会与正负离子形成水合离子，并在正负离子间按电场方向作定向排列，这就大大减弱了离子间的引力(图 2-17)。离子间的作用力可用 $F=\frac{q_1q_2}{r^2D}$ 表示，q_1 和 q_2 代表离子电荷量，r 代表相互间的距离，D 为介电常数。真空的介电常数为 1，水的介电常数为 80(20℃)。这表明离子键在水中的强度约为真空中的 1/80。在细胞这样的含水环境中，离子键的强度相当于氢键，约为 20 kJ/mol。上式显示离子键的强度会随介质的介电常数的不同而异。

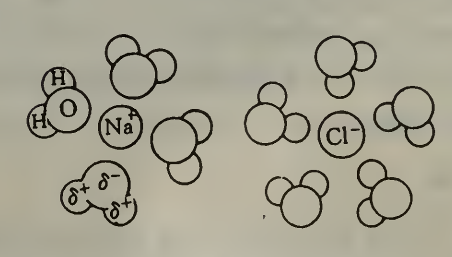


图 2-17　水合离子间的作用

四、短　程　力

短程力(short-range force)是一种当原子或基团接近到很短距离时才明显出现的作用力。短程力的强度约为 5 kJ/mol 左右，并随脱离 r 的增加按 $1/r^4 \sim 1/r^6$ 而迅速衰减。

短程力可存在于离子和分子间及分子和分子间(图 2-18)。

当离子与分子接近时，相互间即逐渐产生静电作用。离子可以与极性分子交互作用，如离子的水合即是。离子也可与非极性分子交互作用，即当它们相互接近时，离子的存在就使非极性分子产生诱导偶极，从而产生微弱的静电引力。

分子和分子间也可发生交互作用，这种交互作用可以发生在偶极和偶极间，偶极与诱导偶极间及诱导偶极间，这些作用力统称为范德华(van der Waals)力。

这种交互作用的强度虽然很弱，但由于生物分子主要都是些大分子，它们的组成原子很多，因此，当它们接近时，这种交互作用在总体上就会变得不可忽视。

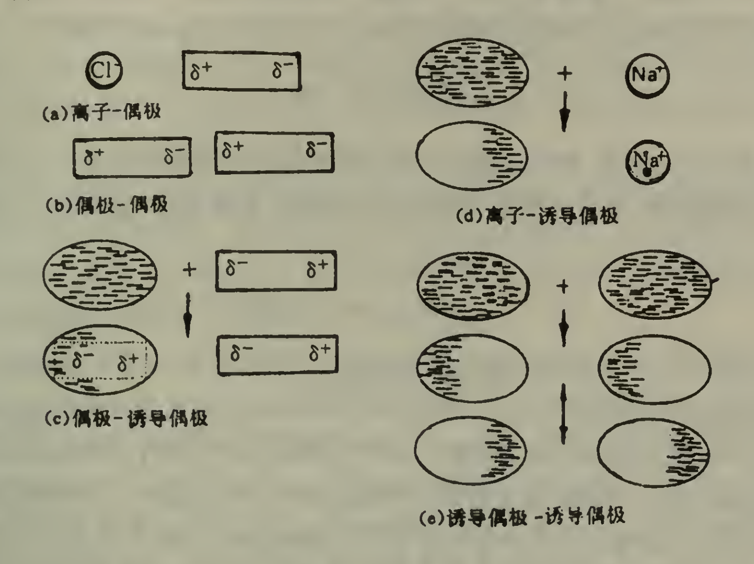


图 2-18　短程力

五、疏水交互作用

当油与水混合时，它们由于不能互溶而分层。一种物质是否溶于水与其分子的极性有关，现以脂肪酸为例来讨论这一问题。在醋酸离子中，甲基虽是非极性的，但它很小，强极性的羧基使整个分子成为水溶性。但在硬脂酸钠中，由于其烃基远较甲基为大，故同样强极性的羧基在分子中所占的比重大为降低，结果就使硬脂酸钠成为两亲化合物，既亲脂又亲水。两亲化合物的分子在水中趋向于在水面形成单分子层，或是在水中形成胶粒(图 2-19)。这种使两亲分子的非极性基(疏水基)避开水而相互簇聚的作用称为疏水交互作用(hydrophobic interaction)。

疏水交互作用有时也称为疏水键，但后一称呼并不合适而容易误导，因为两亲分子疏水基团的簇聚并不是相互间存在某种键而导致的。胶粒中，两亲分子的疏水基间可以存在范德华力，但它不是疏水基簇聚的起因而是其结果。

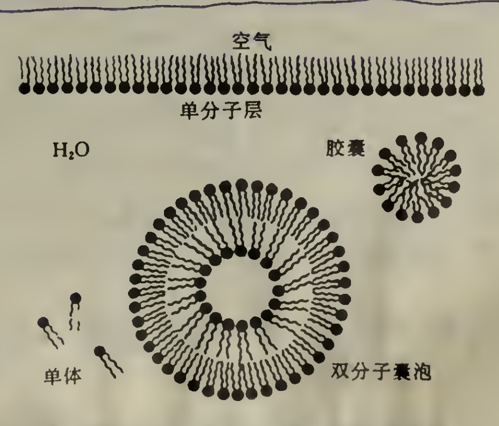


图 2-19　两亲分子形成的单分子层和胶粒

当两亲分子与水接触时，它的疏水基难于与极性的水分子交互作用，迫使周围的水分子间的氢键有所增加。亦即增加了疏水面上水分子的有序排列。液态水的分子是通过氢键在一定程序上缔合的，但氢键会在水分子间转移，水分子并未形成晶格，仍可或多或少地自由运动。水分子间氢键的增加意味着分子运动自由度的降低，即水分子进一步走向有序。在这一过程中所消耗的能量超过形成更多氢键所释放的能量，故这一状态在能量上是不利的。当两亲分子的

疏水基相互簇聚时，上述能量上的不利状态就得到解除，故疏水交互作用是一个主要由水分子混乱度改变所驱动的过程（这将在第五节中作进一步的讨论）。

许多生物分子是两亲的，疏水交互作用对细胞的结构和功能都很重要。

六、配　价　键

一些金属离子能与含氮、氧的基团间形成一种特殊类型的共价结合。$Cu(OH)_2$ 是一种淡绿色的不溶于水的物质，加入浓氨水后即形成一种深蓝色的溶液，这是由于形成了铜氨络离子。

$$Cu^{2+} + 4:NH_3 \longrightarrow [Cu(NH_3)_4]^{2+}$$

氨分子中的氮各以一个单键与氢结合外还具有一孤对电子，后者可与铜离子形成配价键。配价键的本质也是电子对的共享，与一般共价键不同的是共享电子对全由键合原子的一方提供。由 Cu^{2+} 与 NH_3 形成的离子称为络离子，其中的铜称为中心离子，氨称为配位体，含这种结构的化合物称为络合物。许多过渡金属离子都能形成络合物。此外，上节提到的金属离子的水合物中，水的氧原子与金属离子间虽存在着离子及偶极间的作用力，但也存在配价键的成分。配价键的强度约与其它共价键相当，常在 250 kJ/mol 上下；在细胞的含水介质中，由于水具有形成配价键的能力，因此，它实际上的强度较上述数字为低。

络合物中，同一分子具有2个或更多的配位基的化合物称为螯合物（chelate），它们广泛存在于细胞中。卟啉是许多蛋白质的辅基，它也是一种螯合剂（chelating agent），能与金属离子螯合，血红素就是一个例子（图 2-20）。血红素为一扁平的分子，其中由 18 个原子构成的共轭体系处于同一平面中，Fe^{2+} 与 4 个氮络合。在血红蛋白中，血红素在平面的一侧与组氨酸侧链的咪唑基氮络合，在另一侧与氧络合。血红蛋白在呼吸过程中起着运输氧的功能。除了血红素外，细胞色素、过氧化氢酶、维生素 B_{12} 等也都含卟啉辅基，只是所螯合的金属离子不尽相同。

图 2-20　血红素的结构

络合物在生物系统中具有很高的重要性。

第五节　生命系统和热力学

生物是物质进化的产物，细胞是由各种生物分子组成的，但无生命物质和有生命物质具有不同的特征，前者自发地由复杂向简单、由存在差异向均匀方向发展，后者则由简单向复杂、由低级向高级发展。两者形成了鲜明的对比。既然生命是遵循物理和化学的基本规律的，那么我们怎样来理解有生命和无生命物质的这一区别呢？

一、三种热力学系统

细胞是机体的组成单位，在生命过程中它不断地与环境进行着物质和能量的交换。热力学是一门研究热和其它形式能量间的关系和预测物质的物理和化学过程方向和限度的学科。在热力学研究中，我们把给予特殊考虑的宇宙一部分称为系统，而宇宙的其余部分称为环境。细胞和机体就是一个系统，它也遵循热力学的规律。

系统可以分为三种：孤立系统、封闭系统和开放系统。孤立系统与其环境间没有物质和能量的转移，整个宇宙可以看做是一孤立系统。封闭系统与环境间虽无物质的转移，但可发生热、功和辐射等的转移。处于一封闭容器中的物质构成了一个封闭系统。开放系统是一种能与环境发生物质和能量转移的系统。细胞即是一开放系统。

二、能的转化和守恒

能量概念最早是在机械现象研究中建立起来的。例如，斜坡上物体的下滑会做功，后者是物体的势能转化来的。以后，热机发明了，人们在改进热机效率的努力中，进一步认识到了能的一些特性，热力学随之形成。热力学专门讨论热和能的其它形态间的关系，并用以描述和预测宏观物体的物理和化学过程。热力学所讨论的物质性质包括压强(P)，体积(V)，温度(T)，热(q)，功(w)和内能(E)。

系统和环境间的能量交换可以采取多种形式，例如，热、光、机械功、电功等。现在先讨论封闭系统与环境间热(q)和功(w)交换的定量关系。当系统从环境吸收热时 q 定为正，放热时 q 为负；向环境做功时，w 为正，反之为负。在这一过程中，能的形态虽然改变了但量是守恒的：

$$\Delta E = E_{终} - E_{始} = q - w$$

这就是热力学第一定律，式中的 $E_{始}$ 表示过程开始时的能量，$E_{终}$ 表示过程终止时的能量，ΔE 表示两者之差。E 代表系统的内能，包括分子的平动能、转动能、振动能和电子能。但必须指出，这些能量的变化是量子化的，即分为独立的、不连续的能级。例如，氢原子的核外电子，一般情况下处于最低能级(即基态)中，当它吸收到 1.6×10^{-8}J 能量时即跃入第二能级中，如吸收1.9×10^{-18}J 和 2.1×10^{-18}J 能量时就分别跃入第三、第四能级中。能级只采取特定的值。

在分子中还存在其它形式的运动，一个双原子分子中的两个原子可以振动，即相互接近或远离。振动能也是量子化的，数量级为 $10^{-20}\sim10^{-19}$J，数值与分子的结构有关。这比电子能级的间隔小。分子还能转动，能级间差别的数量级为 10^{-21}J。最后，分子还能在空间平移，它的能级间隔更小，为 10^{-35}J。在这几种能量中，电子能对内能的贡献最大，然后依次为振动能、转动能和平动能。

系统的内能为一状态函数，即它依赖于其状态，而与到达该状态的途径无关，故过程的内能差可表达为：

$$\Delta E = E_{终} - E_{始}$$

焓(H，enthalpy)是热力学的另一个状态函数。一个化合物，一种理想气体的焓变为系统的内能加上压强和体积变化的乘积，即焓分子自身的能和分子的相互作用，因为分子间存在着并不涉及化学反应的相互作用。在一个过程中，焓的变化用 ΔH 表示，焓与内能的关系可用下式

表示：

$$\Delta H=\Delta E+\Delta PV$$

当焓的变化很微小时，

$$dH=d(E+PV)=dE+PdV+VdP=dq-dw+PdV+VdP$$

对一个只作 PdV 功的系统，$dw=PdV$，故

$$dH=dq+VdP$$

由于系统相应的内能变化为 $dE=dq-PdV$，故当系统体积不变时，

$$dE=dq \qquad 或\ \Delta E=q_v$$

如过程发生于恒压下，

$$dH=dq\ 或\ \Delta H=q_P$$

上式中的 q_v 和 q_P 分别表示恒容和恒压下的热量。这表明，当系统体积恒定时，过程的内能变化等于它所吸收的热，而过程的焓可由恒压下系统所吸收的热测得，由于大多数的生物过程的压强和体积变化都很小，故 ΔH 和 ΔE 的差异实质上微不足道。这对研究生命活动有很大方便。例如，测定生物组织中葡萄糖完全氧化时焓的变化是一个很困难的问题，这是因为它要经历许多代谢反应，逐步测定无疑难于做到，何况组织中同时还有许多与葡萄糖氧化无关的反应在进行着。既然生化反应常在恒压下进行，焓又是一个状态函数，只与产物和反应物的能量差有关，并不涉及反应的机理与速度，故葡萄糖氧化成二氧化碳和水时的焓变就可通过采用等压量热器的方法来测得，因为在这一方法中葡萄糖完全转化为二氧化碳和水。

三、过程的自发性

在生命活动的研究中，我们常希望了解一个过程的自发性(spontaneity)。但热力学第一定律并没有为我们提供判别能否自发进行的标准。

人们一般常会认为，放热过程总是自发的，因为它代表系统由高能量向低能量水平转变，ATP 的水解确是如此，但情况并不都尽然。如水溶性蛋白溶液与 Cu^{2+} 作用会自发生成沉淀，可是这一过程是吸热的。已如前述，机体的生化过程的内能变化一般等于其热能变化，但热能变化并不能成为判别过程是否自发的标准。因为一个过程发生前后的能量变化由两部分组成：一部分是分子的键能和构象变化等所涉及的能，它在常温、常压下可以做功，称为自由能(free energy，F)；另一部分束缚于系统的混乱度(即系统组分的自由度)，并与温度有关，称为熵(entropy，S)。它在常温、常压下不能被利用。只有自由能才是判别一个过程能否自发进行的标准。自由能也称为 Gibbs 函数(G)。一个过程的自由能变化(ΔG)可以表示为：

$$\Delta G=\Delta H-T\Delta S$$

现在我们来逐步讨论这一问题。在下一反应中

$$aA+bB \rightleftharpoons cC+dD$$

反应开始时，A 和 B 逐步转变为 C 和 D，直至建立平衡为止，这一过程的自由能变化为：

$$\Delta G=\Delta G^{\circ}+RT\ln\left(\frac{[C]^c[D]^d}{[A]^a[B]^b}\right)$$

上式中的 ΔG° 为标准自由能变化，即是 25℃，1 大气压和反应物、产物都是单位活度时的自由

能变化，当反应建立平衡时，$\Delta G=0$，故：

$$\Delta G^{\circ}=-RT\ln(\frac{[C]^c[D]^d}{[A]^a[B]^b})=-RT\ln K_{平衡}$$

通过这一式子，我们可以由 ΔG°计算出反应的平衡常数，或由平衡常数计算出 ΔG°。表 2-5 列出了 $K_{平衡}$和 ΔG°间的数值关系。由表 2-5 可以看出，25℃时 $K_{平衡}$每变化 10 倍，ΔG°变化 5.7 kJ/mol。ΔG°的这一变化还小于氢键键能的一半。

表 2-5　25 ℃时 $K_{平衡}$与 ΔG°间的关系

$K_{平衡}$	ΔG°(kJ/mol)
10^6	34.3
10^4	22.8
10^2	11.4
10^1	5.7
10^0	0.0
10^{-1}	−5.7
10^{-2}	−11.4
10^{-4}	−22.8
10^{-6}	−34.3

从表 2-5 看出，过程的自由能变化由系统的焓变和熵变共同决定的这一关系，对分子生物学有着重要意义。我们先从几个具体例子讨论起。

我们知道，葡萄糖酸常以环型内酯形式存在，它是一种稳定的形式，但羧酸和醇分子间生成的酯并不很稳定，现在我们就来分析这两种情况的差别。羧酸和醇分子间酯化反应的 K 接近 1。

$$CH_3CH_2CH_2\overset{\overset{\large O}{\|}}{C}—OH\ +CH_3CH_2OH \underset{}{\overset{K\simeq 1}{\rightleftharpoons}} CH_3CH_2CH_2\overset{\overset{\large O}{\|}}{C}—O—CH_2CH_3\ +H_2O$$

而醇酸的分子内酯化的 $K_1\gg1$。

$$\begin{matrix}CH_2CH_2COOH\\ |\\ CH_2CH_2OH\end{matrix} \overset{K_1\gg1}{\rightleftharpoons} \begin{matrix}CH_2CH_2—C{=}O\\ |\qquad\quad\ |\\ CH_2CH_2—O\end{matrix}\ +H_2O$$

表 2-5 指出，ΔG°与 $K_{平衡}$间存在着紧密的依赖关系。上述两反应间平衡常数的不同显然是 ΔG°的差异造成的。现在我们再来考察两反应的焓变和熵变。两个反应中都有酯键形成，故焓变应该相同且都为负值。但两个反应的熵变不同。在第一个反应中，羧酸和醇分别为两个分子，反应后羧基和羟基转入同一分子中，分子运动受到的影响较大。羟酸的羧基和羟基本在同一分子中，故反应后分子运动受到的影响小，这使第一个反应中熵的改变大于第二个反应。

分子内反应　$\Delta G_1=\Delta H-T\Delta S_1$

分子间反应　$\Delta G_2=\Delta H-T\Delta S_2$

结果第二个反应 ΔG_2 更负，反应更趋完成，环型内酯更为稳定。

再讨论一个例子。前面提到过的 Cu^{2+} 与水溶性蛋白形成沉淀的过程是自发进行的。但测定的 ΔH 为 12 kJ/mol，这不利于过程的自发进行。但我们还需考虑反应物和产物分子运动自由度的变化。蛋白质和 Cu^{2+} 的结合虽然使运动自由度降低，但两者在结合前都是水合的，一旦结合，原来被束缚于水合物中的水分子被释放，即熵增加了，这有利于过程的自发进行。由于熵增大于焓增，故 ΔG 为负值，反应能自发进行。这时我们是把溶液作为一个系统来考虑的。

疏水交互作用是细胞结构中的一种重要作用力。现在我们再从热力学的角度作进一步的

分析。一个疏水基团插入水分子间产生了两方面的效应：它一方面迫使周围水分子间形成更多的氢键，ΔH 有利于过程的进行；另一方面周围水分子运动的自由度降低了，ΔS 不利于过程的进行，并且超过了 ΔH 的效应，使 ΔG 不利于过程的自发进行（见表 2-5）。但当疏水基团簇聚在一起时，上述情况就逆转，故疏水交互作用的出现是一种熵驱动的过程。

一个过程的自发性是由过程的焓变和熵变共同决定的，现在我们把这三者之间的关系概括如下：当 ΔG 为负，过程能自发进行；ΔG 为正，过程不能自发进行；$\Delta G=0$，过程建立平衡。ΔG 与 ΔH 和 ΔS 间的关系则可归纳为表 2-6。

表 2-6 过程的自发性与 ΔH，ΔS 的关系

ΔH	ΔS	$\Delta G=\Delta H-T\Delta S$
−	+	焓变和熵变都有利于过程的自发进行，温度对自发性无影响
−	−	焓变有利，熵变不利于过程的进行，过程只有当温度低于 $T=\frac{\Delta H}{\Delta S}$ 时才自发进行
+	+	焓变不利，熵变有利于过程的进行，过程只有当温度高于 $T=\frac{\Delta H}{\Delta S}$ 时才自发进行
+	−	焓变和熵变都不利于过程的进行，过程在任何温度下都不能自发进行

四、熵和热力学第二定律

过程的方向和限度是我们讨论一个系统时的基本问题，它们是由系统的焓和熵共同决定的。焓的意义已讨论得比较充分，熵则还有进一步讨论的需要。

为了便于解释熵的意义，我们先从一个简单模型讨论起。假设有一个可以平均分隔为二的箱子，其中放置了 4 个完全相同的球（讨论时，为方便起见暂把它们分别标为 A，B，C，D）。在过程开始时，A，B，C，D 都位于左方（L），以后它们随机移动，其中一些球就移向右方（R），表 2-7 列出的就是 4 个球随机移动的可能结果。

表 2-7 4 个球的随机排布

球				在左方的球数	在右方的球数
A	B	C	D		
L	L	L	L	4	0
*	*	*	*		
L	L	L	R	3	1
L	L	R	L	3	1
L	R	L	L	3	1
R	L	L	L	3	1
*	*	*	*		
L	L	R	R	2	2
L	R	L	R	2	2
L	R	R	L	2	2
R	R	L	L	2	2
R	L	R	L	2	2
R	L	L	R	2	2
*	*	*	*		
R	R	R	L	1	3
R	R	L	R	1	3
R	L	R	R	1	3
L	R	R	R	1	3
*	*	*	*		
R	R	R	R	0	4

在这一由 4 个球组成的系统中，每一个球可能的排布方式为 2（即或左或右），4 个球共 16 种方式，任一球采取某一特定排布方式的概率为 1/16。但 4 个球实际上没有差别，实际情况是：3 个球出现在左（右）方，1 个球出现在右（左）方的概率为 4/16，2 个球出现在一方的概率为 6/16，4 个球全出现在一方的概率为 1/16。由此可知，出现频率最高的排布为 2 个球在一方，另 2 个球在另一方（图 2-21）。

当箱中的球数增加时，球完全集中在一方的概率就越来越小，如球数增加至 20 时，这一数字仅为 $(\frac{1}{2})^{20}$。图 2-22 表示的是箱中的球数

(N)分别为 4,20,和 100 时的概率分布曲线。通过比较,我们可以看出,当 N 增加时,曲线就越来越窄,这表明它与球在一方的平均数的偏差越小。由此可以引出结论,系统中所有粒子集中于一侧的情况,从能力学言是可能的,从统计学言实际是不存在的。

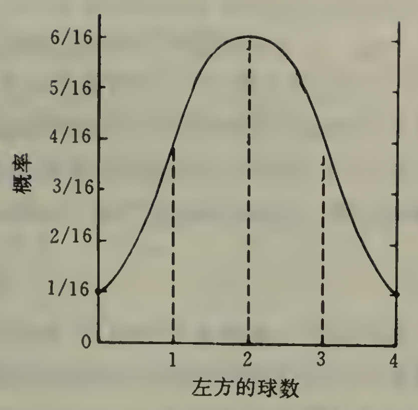


图 2-21 4 个球在箱子两半的概率分布

现在我们再回到前述的模型,并假设它是一个孤立系统,而箱子由一假设的可开启的隔板分隔为二。开始时箱的左方充满 1 mol 理想气体,当隔板开启的瞬间,分子开始采取许多方式排布(即位于左方或右方),随着时间的消逝,分子在箱的左右方的数目趋于均匀,即建立起了平衡。这时宏观上不再表现出变化。或者说,在这一过程中最可能出现的状态即是分布概率最高的,或是最混乱的状态。熵即是系统混乱度的量度。

一个孤立系统总是自发地朝熵增加的方向发展($\Delta S>0$),直至建立平衡,即 $\Delta S=0$。这就是热力学第二定律。

$$\Delta S \geqslant 0$$

由于宇宙不再有环境,故宇宙是一孤立系统,宇宙的熵趋向于一最大值。

熵和内能、焓一样是系统的状态函数,系统的熵变仅决定于其状态。

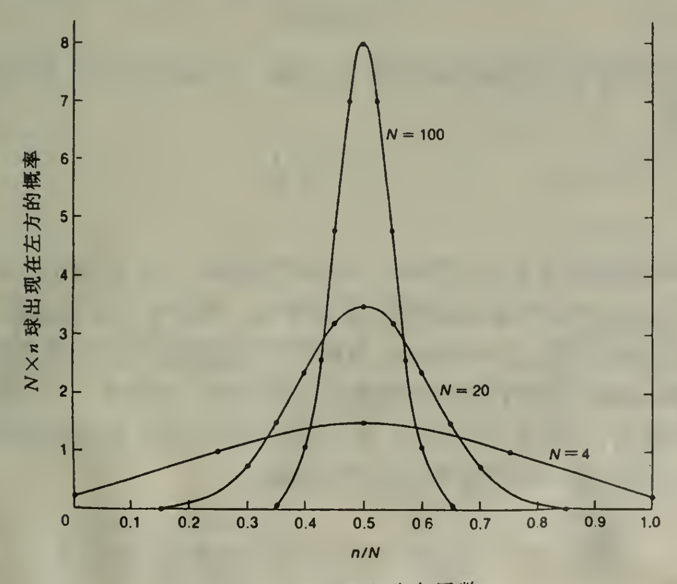


图 2-22 粒子概率分布函数

在前面的讨论中,为了使情况简化从而便于说明问题,我们忽略了粒子排布中的位置差异,现在如以 W 代表特定状态下系统组分的所有可能的排布数,则熵与 W 的关系就可表示为:

$$S=R\ln W$$

R 为气体常数。

现在我们再从分子运动角度来讨论熵。系统的内能包括分子的平动能、转动能、振动能以及电子能。在这些能量中，平动能的能级最密，能级的间隔最小，其次为转动能、振动能和电子能。分子的这些运动中，平动对熵的贡献最大，转动次之，振动最小；对电子言，它们常处于最低能级中，对熵的贡献常为0，这和内能相反。

此外，分子的一些结构因素对熵也是重要的。例如，分子刚性的增加会减少转动和振动能对熵的贡献，这些因素包括双键、三键和环等等。

五、熵与生命

活机体是一高度有序的系统，这种有序性表现在它是由一组特定的生物分子，按照一定的层次和一定的规律构成的。机体各个部分中各类分子的分布是不均匀的，因而功能也不同，但相互间分工精致而又协调，体内的代谢按照一定的程序、一定的方向进行，生物机体是一远离平衡态的开放系统。活机体和无生命物体表现出很大的不同。无生命物体总趋向于无序和均匀，趋向于平衡。生物则不仅能维持其高度的有序性，并且不论在个体发育或是生物进化中都向更高一层的有序性，即向减少熵的方向发展，活机体之所以能如此，因为它是一个开放系统，能与环境发生物质和能量的交换，机体的高度有序性是以增加环境的混乱度(熵)为代价的。

$$\Delta S_{系统}+\Delta S_{环境}=\Delta S_{宇宙}>0$$

已如前述

$$S=k\ln W$$

其中的 W 是系统组分的各种可能的排布数，它反映了系统的无序性，因此，$1/W$ 可用作表示系统有序性的量度。

$$-S=k\ln\frac{1}{W}$$

$-S$ 可称为负熵。

Schrödinger，E. 在他的著名论文“What is life”中指出，一个生命有机体不断地在增加熵，并趋向于一最大值，最后达成平衡，这就是死亡的来临。机体要生存就必须从环境不断汲取负熵。有机体是依赖负熵生存的。Schrödinger 也曾预言生物机体中存在遗传物质，生命的特征之一就是能自我复制，他指出机体内的遗传物质具有晶体的一些特性。以后发现的DNA 证实了细胞中存在遗传物质的预言。Schrödinger 的论点对我们认识生命现象无疑具有重要意义，但如何全面、具体地掌握生命的本质看来还有漫长的道路。

第六节　生物大分子的构象及其组装

一、核酸、蛋白质和遗传信息

核酸(DNA 和 RNA)和蛋白质是机体的两类最基本的生物大分子。它们都属信息分子，但DNA 是遗传信息的原初载体，蛋白质是遗传信息的体现者。或者说，DNA 在机体中代表信息，蛋白质代表功能，但两者是相互依存的。蛋白质的功能是遗传信息规定的，而遗传信息的表达又依赖蛋白质的参与。

在 DNA 和 mRNA 间，遗传信息的传递是通过碱基互补实现的；在 mRNA 和多肽链间，它是由 tRNA 作为中介实现的，使 mRNA 的核苷酸顺序转变为氨基酸顺序。

DNA 是由 4 种核苷酸组成的，核苷酸的排列顺序代表了核酸的语言。蛋白质是由 20 种 α-氨基酸组成的，氨基酸的排列顺序代表了蛋白质的语言，两者存在着对应关系（图 2-23）。多肽链的氨基酸顺序并不能直接表现出功能，功能只是多肽链折叠成特定的三维结构后才出现的，但多肽链的氨基酸顺序包含了它折叠全部信息。Anfinsen 的著名实验证明了这一点。变性使蛋白质丧失其活力，这是它的三维结构受到破坏的结果。Anfinsen 先用尿素和巯基乙醇（$HOCH_2CH_2SH$）处理牛胰核糖核酸酶，结果酶发生变性，原有的 4 个二硫键还原成 8 个巯基。然后用透析法除去尿素和巯基乙醇，酶的活力由于复性而恢复。酶分子中的 8 个巯基可组合成 105 种二硫键，酶活力的全部恢复表明重新形成的二硫键位于原来的位置上，即天然三维结构得到重新建立，它是多肽链自发折叠的结果（图 2-24）。

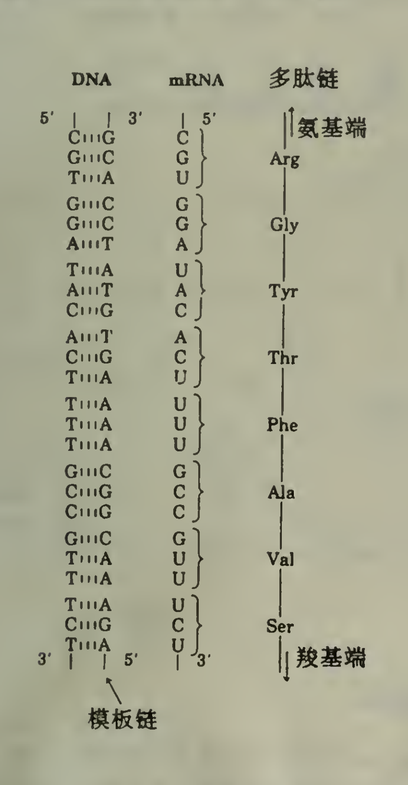


图 2-23 核苷酸和氨基酸顺序间的对应关系

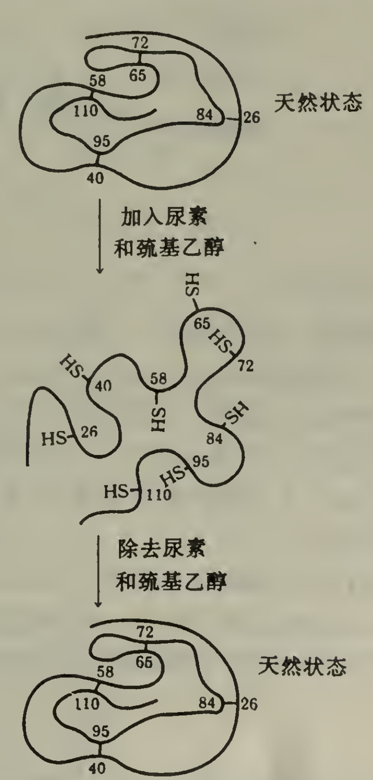


图 2-24 牛胰核糖核酸酶的变性和复性

多肽链是怎样折叠成三维结构的？

二、多肽链的共价结构

蛋白质的结构可分为多个层次，即一级结构（primary structure），二级结构（secondary structure），三级结构（tertiary structure）和四级结构（quaternary structure）（图 2-25）。

多肽链是蛋白质结构的基础，它是由 α-氨基酸按一定的顺序通过肽键连接成的。多肽链的氨基酸顺序称为蛋白质的一级结构。

肽键是一个氨基酸的 α-羧基与另一氨基酸的 α-氨基间失水缩合成的。

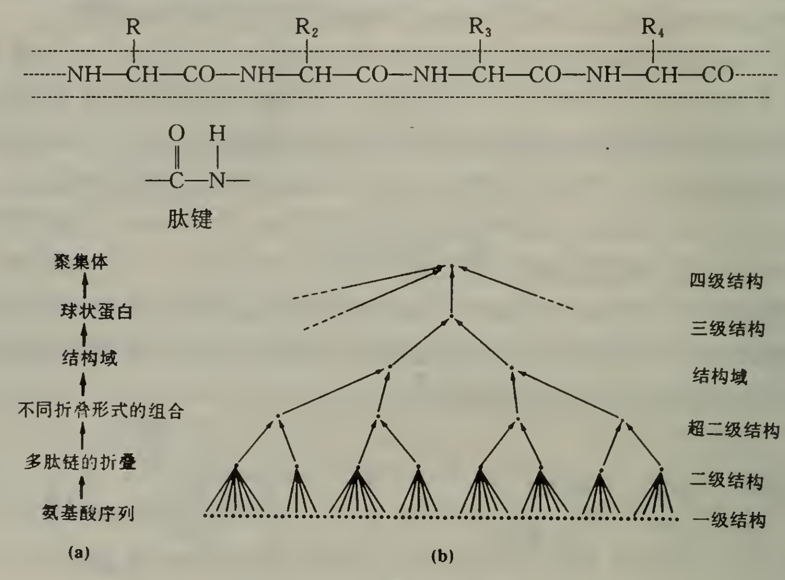


图 2-25 球状蛋白的结构层次

由于共轭效应，$>C{=}O$ 和 $>NH$ 间的单键带有双键性质，在一般情况下不能转动。肽键是一个具一定刚性的接近平面的结构（图 2-26），因此，多肽链的主链可以看做是由 α-碳和肽键交替连接成的。不同多肽链的主链完全相同，不同的是主链上 R 基的种类和排布，它是由遗传信息规定的。

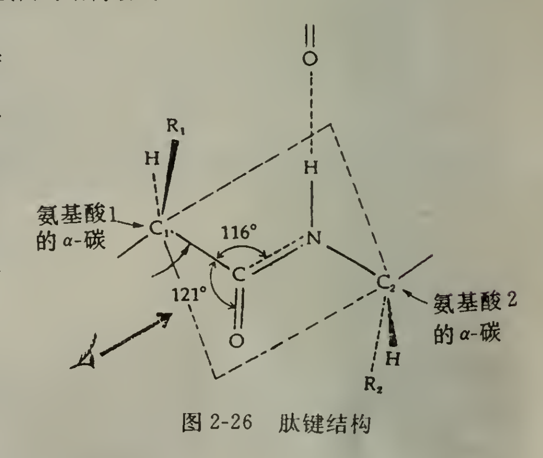


图 2-26 肽键结构

蛋白质氨基酸共有 20 种，它们的差别在于 α-碳上连接的基团（表 2-8）。这些基团决定了多肽链可采取的构象和所表现出的功能。

三、多肽链的构象

在多肽链的主链中，由于肽链为一整体，故能转动的只能是 α 碳上的两个单键 C_α—N 和 C_α—C。由于多肽链含数以百计的氨基酸残基，在理论上由绕单键旋转产生的构象应是一天文数字。但实际上，多肽链可采取的构象是有限的，因为多肽链主链的组成原子间、主链原子和侧链间都可发生干扰，这些干扰来自原子或基团的大小和形状，它们的极性和所带的电荷等。这些限制因素的存在使有些构象是不允许的，有些构象虽然可能但是不稳定的，只有一定的构象是化学上允许和能力学上有利的。这些限制因素使 C_α—N 和 C_α—C 只能在一定的角度内转动，也就是说多肽链的构象可以通过 C_α—N 和 C_α—C 转动的角度 Φ 和 ψ 来表述。氨基酸 R 基对多肽链构象的制约就是遗传信息决定蛋白质结构和功能的实质所在。我们如以 Φ 对 ψ 做图就得到多肽链的构象图或称为 Ramachandran 图（图 2-27），其中实线框标出的是 Φ 和 ψ 的允许组合，网点区表示的是虽有空间障碍，但实际仍可存在的构象。这些允许的 Φ，ψ 组合规定了

表 2-8 蛋白氨基酸的侧链和符号

R的结构	R的性质	名称	符号	R的结构	R的性质	名称	符号
H—	位阻最小	甘氨酸	Gly,G	CH_3—CH(OH)—	极性	苏氨酸	Thr,T
CH_3—	非极性	丙氨酸	Ala,A	$HOOCCH_2$—	酸性	天冬氨酸	Asp,D
$(CH_3)_2$CH—	非极性	缬氨酸	Val,V	$HOOC(CH_2)_2$—	酸性	谷氨酸	Glu,E
$(CH_3)_2CHCH_2$—	非极性	亮氨酸	Leu,L	H_2N—C(=O)—CH_2—	极性	天冬酰胺	Asn,N
$CH_3CH_2CH(CH_3)$—	非极性	异亮氨酸	Ile,I	H_2N—C(=O)—$(CH_2)_2$—	极性	谷氨酰胺	Gln,Q
$H_2C(CH_2—)(CH_2—)$	非极性	脯氨酸	Pro,P	HS—CH_2—	弱酸性 形成二硫键	半胱氨酸	Cys,C
C_6H_5—CH_2—	非极性	苯丙氨酸	Phe,F	$CH_3S(CH_2)_2$—	甲基供体	甲硫氨酸	Met,M
HO—C_6H_4—CH_2—	弱酸性	酪氨酸	Tyr,O	咪唑基—CH_2— (HN, N)	极性	组氨酸	His,H
吲哚基—CH_2— (N, H)	非极性	色氨酸	Trp,W	H_2N—$(CH_2)_4$—	碱性	赖氨酸	Lys,K
HO—CH_2—	极性	丝氨酸	Ser,S	H_2N^+=C(H_2N)—$NH(CH_2)_3$—	碱性	精氨酸	Arg,R

注：氨基酸的符号有两种．一种由 3 个字母组成；另一种仅含 1 个字母。

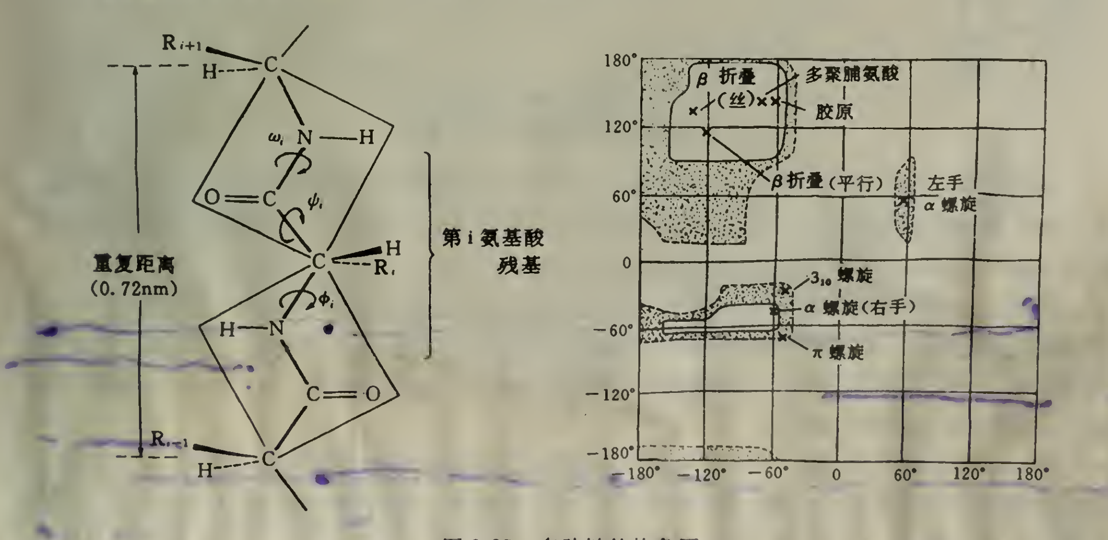


图 2-27 多肽链的构象图

蛋白质的二级结构。蛋白质的二级结构有α-螺旋、β-折叠和β-转角(图 2-28)。α-螺旋相当于$\Phi = -60°$，$\psi = -45 \sim 50°$。它的主链呈螺旋走向，螺旋的半径为 0.23 nm，每匝螺旋含 3.6 氨基酸残基，每一残基的轴向距离为 0.15 nm，故每 0.54 nm 重复一次。α-螺旋中，上下两匝主链的 >NH 和 >C=O 间都形成了氢键；在螺旋内部存在范德华尔力，R 则突出于螺旋的外侧。α-螺旋是一种能力学上稳定的构象。β-折叠中，多肽链为伸展的稍呈扭曲的锯齿状，相邻氨基酸残基的 R 基交替指向锯齿形结构的上下方。β-折叠的Φ小于±180°，并未完全伸展。β-折叠中相邻多肽链的 >NH 和 >C=O 间也存在着氢键。β-折叠有平行和反平行的两种(指多肽链的走向)，图中表示的是反平行的。这两种β-折叠以反平行的较为稳定。β-转角是一种位于多肽链转折处的结构。β-转角含 4 个氨基酸残基，它为第一个氨基酸残基的 >C=O 与第四残基的 >NH 间的氢键所稳定。β-转角有 Ⅰ 型和 Ⅱ 型两种，它们的主要区别在于第二个肽键的空间取向。甘氨酸和脯氨酸是β-转角中经常出现的两种氨基酸。这是由于甘氨酸的α-碳上为两个氢，故在减轻空间障碍中表现出很大的灵活性；脯氨酸则相反，它的α-碳参加了环状结构，故而降低了Φ的自由度，有利于对构象的制约。

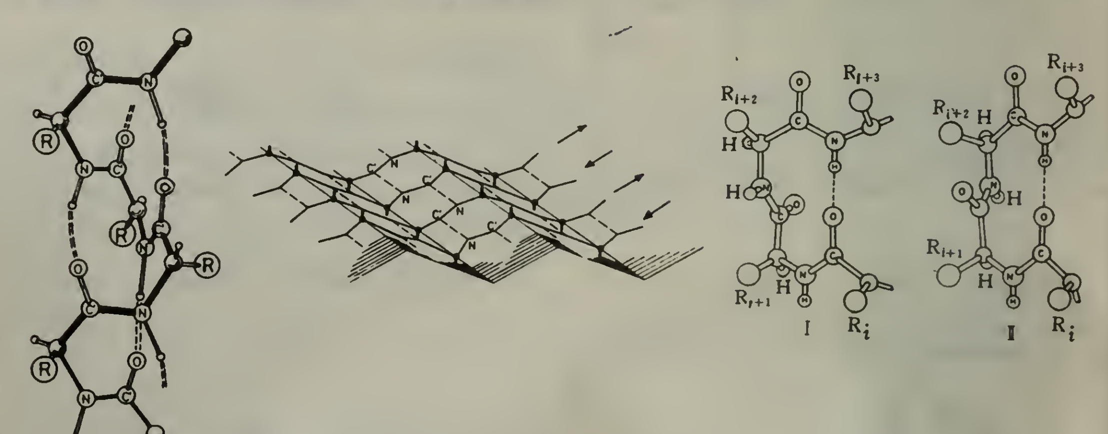


图 2-28　蛋白质的二级结构

左为α-螺旋，中为β-折叠，右为β-转角。

在许多球状蛋白质中，我们常可观察到一些二级结构的组合形式，它们称为超二级结构，超二级结构的经常出现，表明它们在多肽链折叠过程的动力学和能力学上都是有利的。最常见的超二级结构有：$\alpha\alpha$，$\beta\alpha\beta$ 和 $\beta\beta\beta$ 三种图式(图 2-29)。

形成二级结构后，多肽链还可进一步折叠盘曲成三维球状结构，它们具有独立的结构和功能，称为三级结构。蛋白质的这种折叠盘曲是由分布于二级结构表面的基团所驱动和稳定化的，亦即是受一级结构指导的。这种折叠使多肽链形成特定的形状和使一些基团进入合适的位置，从而导致相应功能的出现。

在多肽链折叠形成相应的稳定构象的讨论中，我们不能不提到二硫键的作用，对一些低分子量的蛋白质言尤其如此。核糖核酸酶的变性和复性就是一例，胰岛素为我们提供了另一类型的例子。胰岛素含 51 氨基酸残基，刚合成时为一较大的前体，以后切去一小肽段成为胰岛素

原，后者再切去C链后才成为含A链和B链的胰岛素(图2-30)。胰岛素共含3个二硫键，两个存在于链间，一个存在于A链的两部分间。和核糖核酸酶一样，胰岛素的二硫键一经破坏，它亦立即失活。

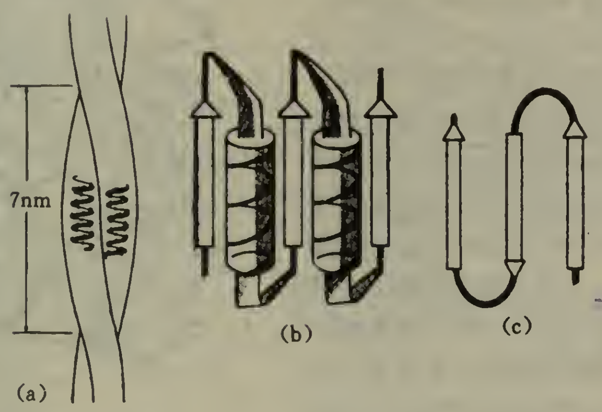


图 2-29 蛋白质的超二级结构

(a)$\alpha\alpha$； (b)$\beta\alpha\beta$； (c)$\beta\beta\beta$。

图 2-30 胰岛素的化学结构(猪胰岛素)

四、蛋白质的三级结构

肌红蛋白三级结构的测定是蛋白质结构研究中的一个突破。肌红蛋白是肌肉中氧的载体，

它是一个含 153 氨基酸残基的单体蛋白，分子量为 16.7 kD。图 2-31 是肌红蛋白三级结构模型，肌红蛋白的三级结构含 8 个 α-螺旋，在图中分别用 A，B，C，D……H 表示（表 2-9）。分子中 R 基的位置也已确定，大部分疏水基都位于分子的内部，而亲水基团则位于分子表面，这一情况说明疏水交互作用在多肽链的折叠中起着重要作用，是折叠的驱动力。肌红蛋白的分子极为紧密，其内部仅能容纳 4 个水分子。一个紧密的疏水核心是球状蛋白的典型特征。肌红蛋白的这种紧密结构使短程力得以充分发挥作用。肌红蛋白含有 4 个脯氨酸残基，它的刚性侧链使其难溶于 α-螺旋内部而都位于转折中。肌红蛋白的辅基为血红素，它位于蛋白质部分的隙缝中。血红素为卟啉化合物，其中心的铁一方面与分子平面一侧组氨酸咪唑基中的 N-3 络合，在另一侧则可与氧络合，从而起着贮氧和输氧功能。这一结构上的安排对肌红蛋白是重要的。由于隙缝位于分子表面而便于氧的接近，但隙缝亦提供了一个氧发挥功能的疏水环境，因为在水中 Fe^{2+} 会迅速氧化成 Fe^{3+} 而难以与氧可逆结合。

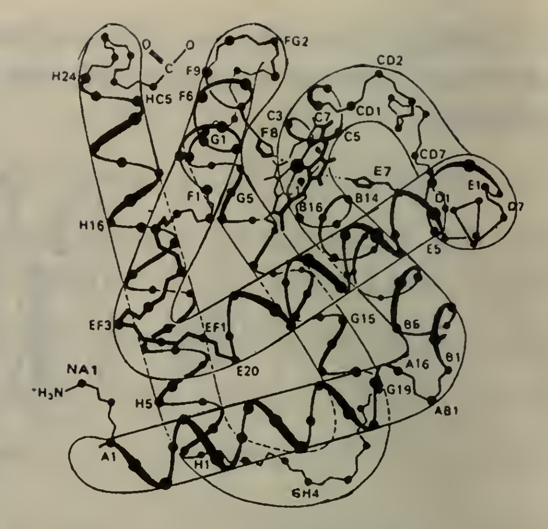


图 2-31　肌红蛋白结构模型

肌红蛋白是以 α-螺旋为主的蛋白质，有些蛋白质兼含两者，也有些蛋白质以 β-结构为主（表 2-10）。

表 2-9　肌红蛋白分子中的 α-螺旋

螺旋长度		非螺旋长度	
A	16	NA	2
B	16	AB	1
C	7		
D	7	CD	8
E_1	10		
E_2	10		
F	10	EF	8
G	19	FG	4
H	26	GH	5
		HC	4
总长度	121		32

表 2-10　一些单链蛋白中 α-螺旋和 β-折叠的含量

蛋白质（残基数）	残基的百分数	
	α-螺旋	β-折叠
肌红蛋白（153）	78	0
细胞色素 c（104）	39	0
溶菌酶（129）	40	12
核糖核酸酶（124）	26	35
凝乳蛋白酶（24）	14	45
羧肽酶（307）	38	17

有些较大蛋白质的多肽链会折叠成相对独立的称为结构域（structural domain）的三维结构。结构域是多肽链的一个部分，但已具有完整球状蛋白质的特征，从这一意义来说它是自主的。它们可以通过有限的蛋白酶解从多肽链上切割下来而不改变其性质。有些蛋白质的结构域相似，有些则具明显差异。现在已证明，结构域常对应于基因中各自的外显子（exon）。结构域的组合就形成了蛋白质的完整三级结构。

磷酸甘油醛脱氢酶为一较大的蛋白质，含334氨基酸残基，形成了两个分别称为NAD^+-结合

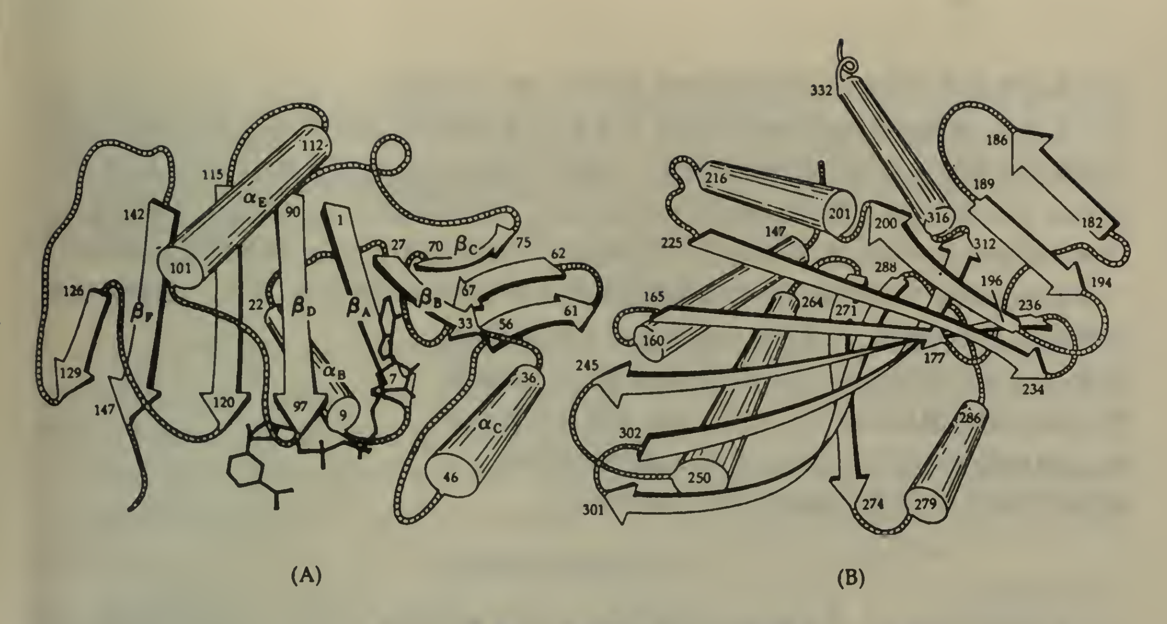


图 2-32　磷酸甘油醛脱氢酶的两个结构域

(A)为 NAD^+-结合结构域；(B)为催化结合结构域。

结构域和催化结构域的部分(图 2-32)。在这两个结构域中，β-折叠都占优势：NAD^+-结合结构域所含的几乎全是平行的 β-结构(由箭头表示)，而催化结构域则含平行和反平行两种 β-结构。两种结构域还都含 α-螺旋，它们位于中心片层结构的两侧。

许多蛋白质，尤其是分子量较大的蛋白质是由几个结构域组成的。

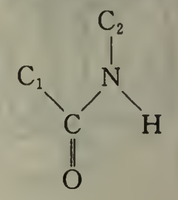
五、多肽链的折叠过程

天然蛋白质是多肽链合成后经折叠而形成的热力学上稳定的构象。多肽链的折叠是一自发过程。在这一过程中，多肽链的共价结构基本上没有变化，起作用的是弱交互作用。多肽链的折叠代表着信息转变为功能的过程，是生命现象的一个重要过程，正受到人们的重视。这一过程现在并不完全清楚，但根据现有的实验证据人们已提出了一些模型。

假设一个多肽，它由 100 个氨基酸残基组成，如根据多肽链中 Φ 和 ψ 可采取的值计算，它形成天然蛋白质至少需 10^5 年。这对一个天然过程是不可想象的。胰岛素的复性也说明了这一状况。胰岛素的分子共含 3 个链间和链内二硫键，如它们全部断裂后再形成，半胱氨酸侧链间的可能组合方式难于计算，亦即很难得到其天然状态。可是中国的科学家在使牛胰岛素的 A 链和 B 链分开，然后重建的实验中，产物的活力竟恢复到 50%。这一结果发表于 1966 年，随后合成了胰岛素的 A 链和 B 链以及获得了天然胰岛素晶体。这是世界上第一个人工合成的蛋白质。这些工作都提示，天然蛋白质虽然是其多肽链折叠所形成的最稳定的一种形式，但多肽链的折叠似乎不是一个不断尝试的过程，而是一个有导引的协同过程。为此，人们现已提出了一些多肽链的折叠模型，大致可以分为二类。一种模型认为多肽链的折叠是逐步进行的，先形成一种稳定的二级结构作为核心，然后二级结构的氨基酸侧链进一步发生交互作用，扩大成天然三维结构；另一种模型提出，多肽链可能由于其疏水侧链的疏水交互作用而突然自发折叠，形成一种含二级结构的紧密状态，最后调整成天然结构。这两种模型看来不是排斥的，有些多肽链的折叠可能以其中之一为主，有些多肽链的折叠兼而有之。在这两种情况下，超二级结构的

形成都可能起着导引作用，而弱键则做最后的热力学上的调整。

活体中，多肽链发生折叠的环境似乎更为复杂，而折叠的效率更高，现已证明，活体内至少有两类蛋白质因子参与了多肽链的折叠，一类蛋白质称为多肽链结合蛋白(polypeptide chain binding protein)或分子伴侣(molecular chaperone)。这些蛋白质中，有些能结合于多肽链以防止侧链间的非专一性缔合，它们导引一些多肽链的折叠和使多个多肽链聚集成更大的结构；另一类蛋白质是酶，它们能通过解除限制折叠的因素来促进多肽链的折叠，例如，二硫键异构酶、脯氨酰顺-反异构酶等。在多肽链的合成过程中肽键可能成为顺式，但顺式较不稳定，与反式比相差 8 kJ/mol，脯氨酰顺-反异构酶能催化其转变为反式。在多肽链的折叠过程中，二硫键可能形成于不适的位置间，二硫键异构酶催化二硫键的可逆形成，以便二硫键形成于正确的位置上。



六、蛋白质的四级结构

有些蛋白质仅含一条多肽链，为单体蛋白，有许多蛋白质含二个或二个以上的亚基，为寡聚蛋白(oligomeric protein)，如血红蛋白由 4 个亚基组成，天冬氨酸转氨甲酰酶由 12 个亚基组成等。在这些蛋白质中，亚基的空间关系和缔合称为四级结构。寡聚蛋白普遍存在于机体中。

肌红蛋白(myoglobin，Mb)和血红蛋白(hemoglobin，Hb)都是氧的携带者，都能与氧可逆地结合。

$$Mb+O_2 \rightleftharpoons Mb \cdot O_2$$

$$Hb+O_2 \rightleftharpoons Hb \cdot O_2$$

但肌红蛋白为单体蛋白，血红蛋白为寡聚蛋白，它们对氧的亲和力有所不同。血红蛋白流经肺部时约有 96% 为氧饱和，但在静脉中仅有 64% 饱和，在这一过程中约有携带的氧的1/3释放出去。可是肌红蛋白在同样的氧分压变化中仅释放出 1%～2% 结合的氧。在功能上的这一差别即在于一个为单体；另一个为寡聚蛋白。血红蛋白的亚基结构为 $\alpha_2\beta_2$，α 亚基含 141 氨基酸残基，β 亚基含 146 氨基酸残基。虽然肌红蛋白与血红蛋白亚基的三级结构十分类似，但血红蛋白的多亚基结构使得各个亚基的氧合作用得以调节。这种调节作用表现为：一个亚基与氧的结合会使其余亚基的氧合更趋容易，这称为协同效应。血红蛋白的协同效应是通过其三维结构上的变化实现的(图 2-33)。肌红蛋白只有一个亚基，故而没有协同效应。蛋白质的四级结构的出现是为了完成更为复杂的生理生化过程。

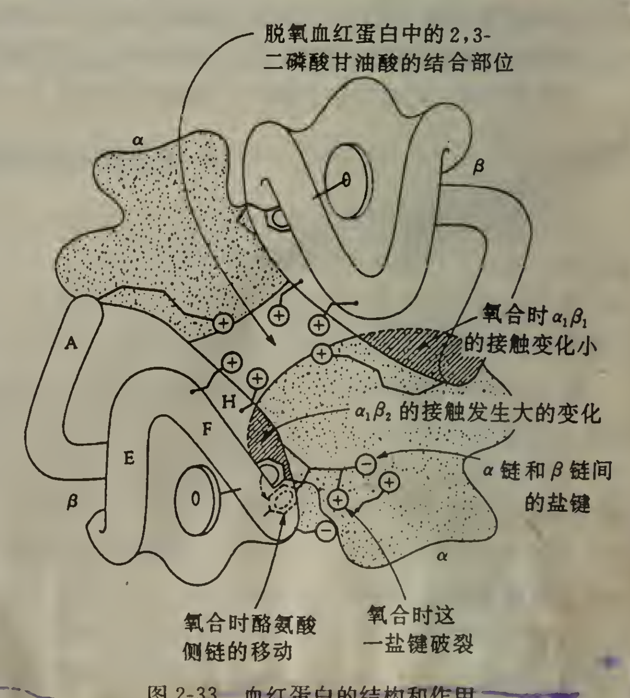


图 2-33 血红蛋白的结构和作用

血红蛋白的 4 个亚基分别位于四面体的顶点上，血红蛋白的 2 个 α 或 2 个 β 亚基都是相

同的，但为了讨论方便起见，故暂标为 α_1,α_2 和 β_1,β_2。在血红蛋白中，α_1/β_1 和 α_2/β_2 间广泛存在着交互作用，这就使 α_1/β_1 和 α_2/β_2 的结合较 α_1/β_2 和 α_2/β_1 更为紧密。这种亚基间交互作用上的差别就使 $\alpha_1/\beta_1,\alpha_2/\beta_2$ 作为整体发生转动，从而引起亚基构象的变化，这就是氧合作用协同效应的结构基础。

七、蛋白质结构的组合性

α-螺旋，β-折叠和 β-转折是多肽链所形成的二级结构，它们是热力学上稳定的构象。二级结构又可组合成不同的超二级结构，很多蛋白质都可看作是一定的超二级结构组合成的。在许多蛋白质结构中，我们都可看到 $\beta\beta$、$\beta\alpha\beta$ 等这样的结构单位。

现已发现，许多超二级结构常具有专一的功能上的意义。例如，螺旋-转折-螺旋这样的超二级结构最初发现存在于 CAP 和 Cro 阻遏物中，两个螺旋几乎成直角，其中之一嵌入 DNA 的大沟中，与 DNA 的专一性序列结合。这一结构单位存在于许多调节蛋白中。

超二级结构和结构域都常与蛋白质特定活力相对应。例如，在以 NAD^+ 为辅酶的磷酸甘油醛脱氢酶、乳酸脱氢酶等分子中都存在一种称为二核苷酸折叠(dinucleotide fold)的结构单位。与此同时，以 ATP 为底物的激酶中也有类似的结构域。这两种结构域，一个与 NAD^+ 结合，一个与 ATP 结合，但它们在结构上具有共同点，即其中心为平行的 β-折叠，两侧为 α-螺旋(参见图 2-32)，NAD^+ 或 ATP 的结合部位则位于 β-结构的羧基端。两者的差异主要在于 α-螺旋的数目和排布。这种状况表明，它们可能是由共同的前体演化来的。

许多蛋白质在功能上的差异似乎也是来自结构域的组合上。例如，脱氢酶都有两种结构域，一个与 NAD^+ 结合；另一个与底物结合，后者的结构在不同的脱氢酶中是不同的。

调节基因表达的阻遏物(repressor)和活化蛋白(activator protein)为蛋白质结构的组合性提供了另一类例子。大肠杆菌的 *lac* 阻遏物和分解代谢基因活化蛋白(CAP)都含多个结构域。CAP 含两个结构域，较大的与 cAMP 结合，较小的识别和结合于 DNA 的特定序列(图 2-34)。CAP 的 cAMP-结合结构域和依赖于 cAMP 的蛋白激酶的调节亚基的氨基酸序列具有明显的同源性，这提示它们可能是由共同的前体演变来的。在 CAP 中，cAMP-结合结构域与 DNA-结合结构域相连接，cAMP 水平的变化直接控制转录水平。依赖于 cAMP 的蛋白激酶为变构酶，它的cAMP-结合结构域所调节的是糖原降解级联系统(cascade)中的第一个酶。CAP

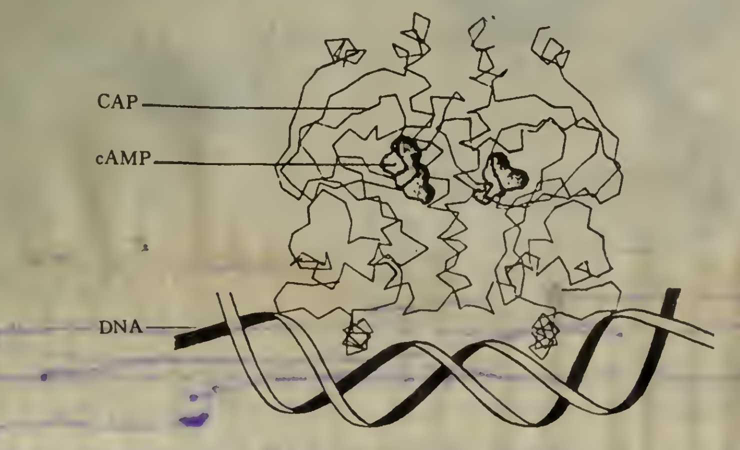


图 2-34　CAP 与 DNA 的结合

的 DNA-结合结构域与其它许多的基因调节蛋白所显示出的结构和氨基酸序列上的类似性，同样表明它们的 DNA-结合结构域可能具有共同的祖先。

lac 阻遏物也具有类似的组合结构。它的小结构域的氨基酸序列与 CAP 的结合于 DNA 的那一部分相关，而它的大结构域则与半乳糖结合蛋白的氨基酸序列类似。这再一次显示，具不同功能的结构域（结合半乳糖和结合特定 DNA 序列的结构域）能组合成另一种蛋白质。

第七节　蛋白质与其它生物分子的交互作用

蛋白质除了通过多肽链的折叠，不同结构单位的组合形成不同的三维结构、产生不同功能外，还能和其它大分子发生交互作用，从而拓展它的结构和功能上的多样性。和蛋白质发生交互作用的生物分子有：多糖、脂质、核酸和蛋白质自身。蛋白质与其它分子之间的交互作用，有的通过共价结合实现，有的通过非共价力实现。

一、蛋白质与其它生物分子的共价结合

糖蛋白（glycoprotein）是蛋白质和寡糖链通过苷键连接成的，寡糖链含有多种单糖，而每一单糖除半缩醛羟基外还含 1 个以上的醇性羟基，故单糖间可通过不同的苷键连接。加上基团空间取向的不同，同一类型苷键又可分为 α 和 β 两型。寡糖链通过其组成单糖的种类，以及结合方式的不同，所形成的多种多样的寡糖链是机体中分子的识别标志。组成寡糖链的单糖有：葡萄糖、半乳糖、甘露糖、鼠李糖、阿拉伯糖、胞壁酸和唾液酸等（图 2-35）。单糖相互间的苷键有 1→2，1→3，1→4，1→6，2→3，2→6 等，半缩醛的构型可以是 α 或 β 型。这种结构上的变化也代表着一种信息。糖蛋白的寡糖链是通过与蛋白质部分中丝氨酸、苏氨酸残基侧链中的羟基形成苷键而连接的。

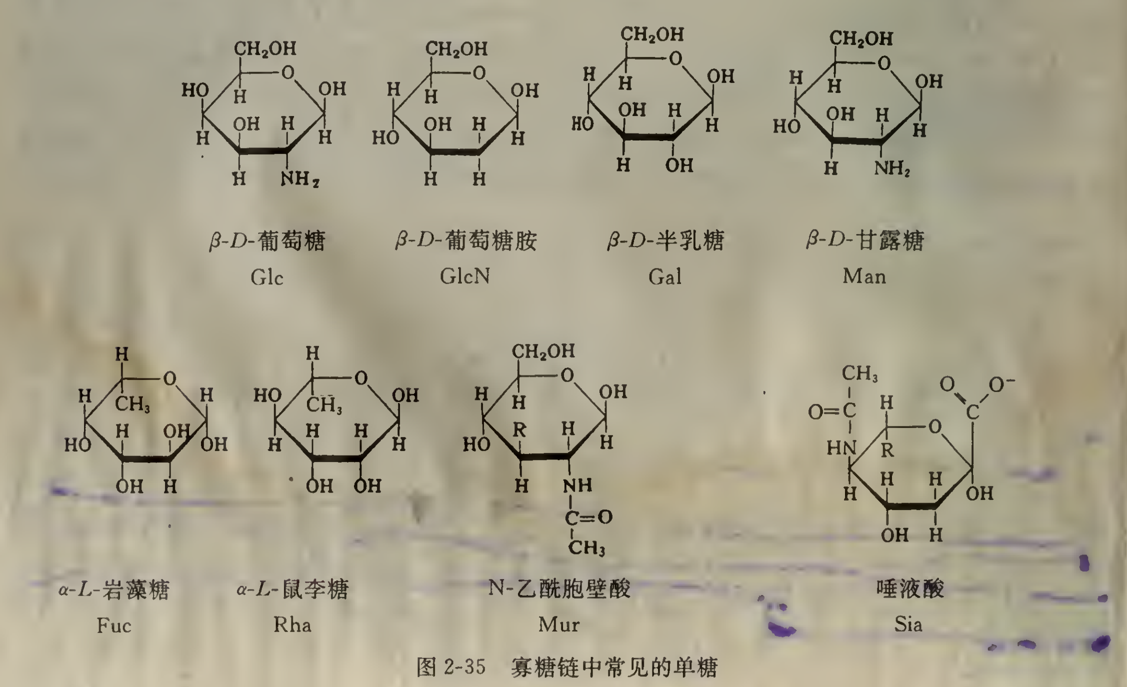


图 2-35　寡糖链中常见的单糖

顺便一提，机体中还有一些蛋白质与脂质共价结合成的蛋白脂(proteolipid)。蛋白脂中的脂质，有些通过酯键与多肽链中的丝氨酸和苏氨酸连接，有些通过硫酯键与半胱氨酸连接。

不论是寡糖链还是共价连接于蛋白质的脂质，都会对后者的结构和性能产生一定的影响。

二、蛋白质形成的超分子复合物

蛋白质自身及蛋白质和其它生物大分子都可能自发聚集成超分子复合物(supramolecular complex)。血红蛋白在尿素的存在下会裂解为两个由 α 和 β 亚基组成的半分子，当尿素去除后，两个半分子又重新缔合成功能性的血红蛋白。血红蛋白裂解为两个半分子是原来存在的弱键被破坏的结果。而半分子的重新缔合是弱键重新形成的结果。生物分子之间的自发缔合称为自组装(self-assembly)。生物分子的自组装涉及到两方面的问题：一是 ΔG 有利于复合物的形成；二是形成复合物的分子具有相应的专一性表面结构。例如，细胞膜和细胞壁都是由许多分子组成的镶嵌结构，现在虽然对膜中不同分子的确切排布还不很清楚，但这些分子是充分接触和存在着交互作用的。血红蛋白的形成是生物分子自组装的一个具体例子，下面对此稍加普遍化。

例如，某一蛋白质分子如具有一个能与自身表面某一区域互补的位点，这样的两个分子就可能自发组装成二聚体。假如蛋白质分子的结合位点互补于结合位点以外的区域，结果它们就可能形成链状结构；如每一亚基采取一定的取向，它们就可能形成环状和螺旋状结构(图 2-36)。肌动蛋白(actin)是一个普遍存在于真核细胞内的球状蛋白，肌动蛋白微丝就是肌动蛋白单体聚集成的。胶原则是一种三螺旋蛋白复合物。蛋白质亚基形成螺旋、环、管或球状结构后，由于亚基间弱键的作用，其稳定性得到进一步的加强，许多细胞结构都是生物分子通过自我组装形成的。

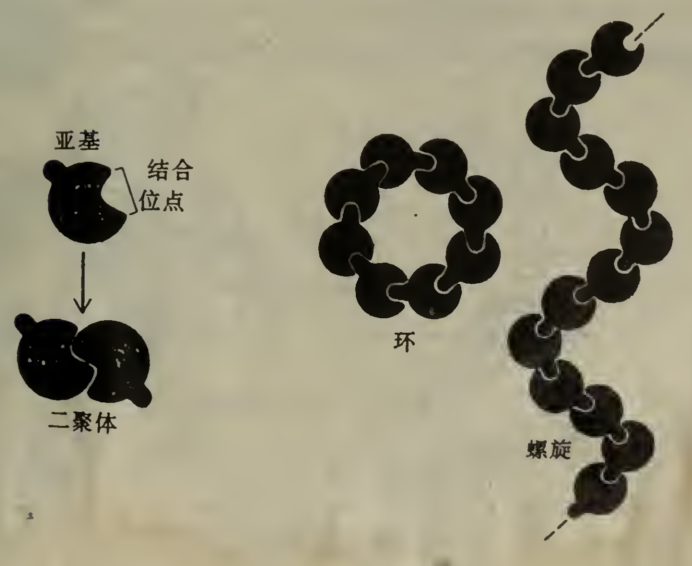


图 2-36　蛋白质单体的聚集

微管(microtubule)是真核细胞的另一种重要蛋白质。它的半径为 25 nm 的中空管状结构，一般由 13 股细丝组成。组成微管的有 α 和 β 两种微管蛋白(tubulin)。α 和 β-微管蛋白先结合成二聚体，然后再组装成微管(图 2-37)。微管和微管蛋白处于动态平衡中，因而会因代谢

条件的改变而趋向于聚集或解体。

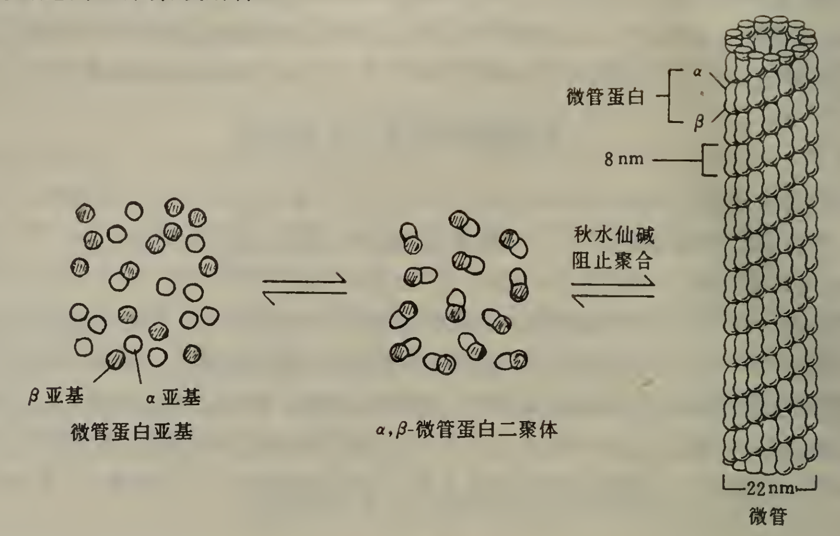


图 2-37 微管的组装

蛋白质还能与核酸形成复合物。病毒粒子就是由 RNA 或 DNA 与蛋白质形成的。例如，烟草花叶病毒(tobacco mosaic virus，TMV)是由 1 个 RNA 分子与 2 130 个相同的蛋白质亚基构成的。病毒粒子的形成仍然离不开组分的前述两种因素。对 TMV 言，各个组分间的作用力，主要的是疏水交互作用。据测定，TMV 组装过程的 ΔH 为 120 kJ/mol，ΔS 为 415 J/mol，K° $\Delta G=\Delta H-T\Delta S$，故 TMV 的组装过程显著受温度的影响(图 2-38)。低于 10℃，ΔG 为正，病毒粒子解体；约 16℃，$\Delta G=0$，两者建立平衡；37℃时，ΔG 为负，过程明显趋向自组装。这是由于温度不同，过程的焓变和熵变在 ΔG 中所占的比重发生变化的结果。

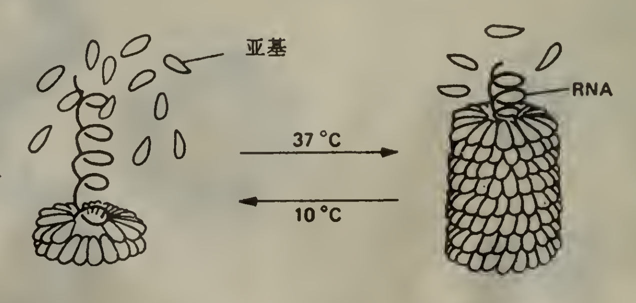


图 2-38 TMV 的自组装

$$\Delta G=\Delta H-T\Delta S$$

37℃ $\Delta G=120\,000-(310)(415)=-8.6\ \mathrm{kJ/mol}$

16℃ $\Delta G=120\,000-(289)(415)\approx 0\ \mathrm{kJ/mol}$

10℃ $\Delta G=120\,000-(283)(415)=+2.5\ \mathrm{kJ/mol}$

除了病毒外，生物机体内蛋白质与核酸的交互作用对生命活动具有十分重要的意义。遗传信息的传递和调控都离不开蛋白质与核酸的交互作用。蛋白质和核酸的交互作用可以分为两个方面：一是核酸与蛋白质形成特定的超分子复合物，如核糖体(ribosome)，核内小核糖核蛋

白(small nuclear ribonucleoprotein, snRNP),信号识别颗粒(signal recognition particle, SRP),延长因子,tRNA 与核糖体形成的三元复合物等。它们为生命过程提供所需的机构。另一种蛋白质与核酸间的交互作用是动态性的,它与基因表达及其调控有关。关于前者的原理已有介绍,现在专就后者作一梗概性的讨论。

DNA 的复制是藉 DNA 聚合酶、RNA 聚合酶及相应的蛋白质因子与 DNA 的结合实现的。作用时,这些蛋白质必须与 DNA 的特定序列结合,这里就有一个蛋白质因子是如何寻找和识别 DNA 特定序列的问题。现在认为,在这一过程中常是蛋白质的 α-螺旋嵌入 DNA 的大沟中,先是其结合部位附近的具正电性区域(由具正电性的氨基酸侧链构成)与 DNA 主链磷酸基的负电荷产生非专一性的静电交互作用。由于两者间的距离较远,作用力因而较弱,故蛋白质能藉热运动而在 DNA 上滑行。当蛋白质移动到靶位点时,由于蛋白质分子特定位点上的氢供体和受体在空间的取向正好与 DNA 靶位点互补,结果氢键得以形成,相互的距离缩短约 1.0～1.5 nm,作用力的强度增强 10^6～10^7 倍,专一性的复合物形成了,酶和蛋白质的功能开始发挥。蛋白质对 DNA 特定序列的这种寻找方式要比在 DNA 上跳动能更迅速地找到靶位点。

Cro 阻遏物与 DNA 的结合是这种交互作用的一个具体例子。X 射线分析等都说明,Cro 阻遏物是以二聚体形式与 DNA 结合,阻遏物的螺旋-转折-螺旋结构突出于其表面,螺旋之一插入大沟中,DNA 稍呈弯曲(图 2-39)。除 Cro 阻遏物外,λ 阻遏物和 CAP 等与 DNA 的结合都涉及螺旋-转折-螺旋这一共同的结构单位。但在 DNA-结合蛋白中,螺旋-转折-螺旋并不是唯一的结构模式。



图 2-39　Cro 阻遏物与 DNA 的交互作用

参 考 文 献

1. 中国科学院生物物理研究所，北京大学化学系，中国科学院有机化学研究所. 结晶胰岛素的合成. 科学通报. 1966. 17:241～271.
2. 冯端，冯步云. 熵. 北京：科学出版社，1992.
3. 原田磐著. 庚镇城译. 生命起源的化学基础. 上海：上海科学技术出版社，1971.
4. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D. Molecular Biology of the Cell. Garland Publishing, Inc. New York: 1989.
5. Dickerson, R. E. Chemical Evolution and the Origin of Life. Sci. Am. 1982. 239(3): 70～86.
6. Klotz, I. M. Energetics in Biochemical Reactions. Academic Press, Inc. New York: 1957.
7. Lehninger, A, L., Nelson. and Cox, M. M. Principles of Biochemistry. Worth Publishers, Inc. New York: 1993.
8. Schrodinger, E. What is Life? Cambridge University Press, 1955.
9. Voet, D. and Voet, J. D. Biochemistry. John Wiley Sons, New York: 1990.
10. Wald, G. Life in the Second and Third Periods; or Why Phosphorus and Sulfur for High-energy Bonds? In: Horizons in Biochemistry. Kasha, M. and Pullman, B. (Editors). Academic Press, New York: 1962.
11. Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. and Weiner, A. M. Molecular Biology of the Gene. The Benjamin/Cummings Publishing Company, California: 1987.

（张玉麟）

第三章　细　　胞

第一节　引　　言

细胞(cell)的研究目前正处于现代生物学研究的中心地位，这是由于细胞是生命的基本单位，细胞在生物的结构上处于承前启后的位置。生物体内由小分子(氨基酸、核苷酸等)组成蛋白质及核酸等生物大分子，再由生物大分子组成超分子结构，如各种细胞器，最终由各种细胞器构成细胞。在多细胞生物中又由各种细胞组成组织(tissue)和器官(organ)，最后形成动物或植物的机体。

生物界是形形色色的，但一切生物都是由细胞组成的，小的生物机体由单个细胞组成，如单细胞藻类(algae)、细菌及原生动物(protist)，它们只能靠细胞的生长和分裂进行发育和繁殖。自从 1665 年虎克(Robert Hooke)用复式显微镜观察到植物木质组织中存在"细胞"之后，施莱登(Matthius Schleiden)于 1838 年证明植物结构由细胞组成，次年施旺(Theodor Schwann)又证实动物结构也由细胞组成，从而产生了细胞概念(cell concept)。德国病理学家维尔荷(Rudoff Virchow)于 1858 年进一步提出了细胞学说(the cell theory)，他认为"细胞是生命的单位(vital unit)"，他说："每个动物是许多生命单位的总和，每个生命单位本身带有生命的全部特性"，现代生物学已经证明了他的观点，细胞的基因组拥有该种生物的全部遗传信息。

在形态上细胞大体上由细胞核、细胞质及细胞膜三部分组成。根据细胞核的区别，生物可以分为原核生物(prokaryote)及真核生物(eukaryote)两大类。原核生物在生物进化上是很原始的，多为单细胞生物，而真核生物由原核生物进化而来，在结构和功能上都日益复杂。真核生物除去核物质外面有一层核模包围以外，其细胞质的结构也比较复杂，其中含有许多细胞器，如线粒体(mitochondria)、高尔基体(Golgi body)、内质网(endoplasmic reticulum)、溶酶体(lysosome)、细胞骨架(cytoskeleton)等，在植物叶细胞中还含有叶绿体(chloroplast)以及液泡(vacuoles)。

第二节　原核细胞与真核细胞的比较

细胞分为原核细胞(prokaryotic cell)和真核细胞(eukaryotic cell)。原核生物是简单的单细胞有机体，它们的核物质没有被一层核膜(nuclear membrane)所包围，核物质散布在细胞质中，未形成明显的细胞核，因此称为原核。细菌(如大肠杆菌 *E. coli*)、蓝绿藻(如鱼腥藻 *Anabaena*)、立克氏体(rickettsia)、支原体(mycoplasma)都属于原核生物。真核生物的核物质是被核膜(nuclear envelope)包围的，形成一个结构非常分明的细胞核(nucleus)，其中还有一个或多个核仁(nucleolus)。真核细胞的细胞质外面有一层质膜(plasma membrane)，成为细胞的外

层。细胞质中含有许多细胞器,这些细胞器也都由膜包围着。在细胞分裂时核膜消失,核物质形成浓缩的染色体(chromosome)。

下面我们将原核细胞与真核细胞做一比较,列于表 3-1。

表 3-1 原核细胞与真核细胞的比较

特 性	原核细胞	真核细胞
尺度	通常很小(1~10 μm)	通常较大(10~100 μm)
基因组	DNA 与非组蛋白结合,基因组存在于类核体中,无膜包围	DNA 与组蛋白及非组蛋白结合,存在于染色体中,染色体存在于具核膜的细胞核中
细胞分裂	二分裂(binary fission),或出芽(budding),无有丝分裂	有丝分裂,含有纺锤体、中心体
膜包围的细胞器	无	线粒体、叶绿体(植物中)、内质网、高尔基体、溶酶体等
营养	吸收,有些进行光合作用(光合细菌)	吸收或进行光合作用(绿色植物)
能量代谢	无线粒体,氧化酶类与质膜结合,多条代谢途径	氧化酶类包装在线粒体中,氧化代谢途径较单一
细胞骨架	无	复杂,有微丝(肌动蛋白丝)、微管及中间纤维系统
细胞内运动	无	胞质流动,内吞作用、胞饮作用、有丝分裂、突触运输等

随着早期原核细胞的进化,产生出各种单细胞的真核生物,如原生动物(protist),其中有些能进行光合作用,如衣藻(*Chlamydomonas*)以后出现了非光合的原生动物,如草履虫(*Paramecium*)及网柄菌(*Dictyostelium*)。单细胞真核生物在地球上是大量存在的,而多细胞的动物、植物和菌类都属于真核生物。原核生物只有数千种,而真核生物则有数百万种之多。

大约 15 亿年前,从原核细胞进化出真核细胞。从原核细胞到真核细胞的进化主要有下列三方面的进化。首先是细胞获得更多的 DNA,见表 3-2。

表 3-2 DNA 含量及基因组的复杂性

	基因组大小 (nt 对)	基因组相对大小 (*E. coli*=1)	DNA 长度 (mm)
病 毒			
SV40	5×10^3	0.00125	0.0017
T7	4×10^4	0.01	0.014
T2	2×10^5	0.05	0.68
原核细胞			
Mycoplasma	3×10^5	0.075	0.10
Bacillus	3×10^6	0.75	1.06
E. coli	4×10^6	1.00	1.36
真菌			
酵母	2×10^7	5	68
动物			
果蝇	2×10^8	50	6.8
鸡	2×10^9	500	680
人类	5×10^9	1 250	1 700
植物			
豌豆	1×10^9	2 250	3 100
小麦	1×10^{11}	30 000	34 000

DNA 与一些特殊的蛋白质(组蛋白等)结合,折叠并由膜包装成染色体,在有丝分裂时可

以平均分配到两个子细胞之中。

其次，由于细胞变大，在细胞内产生一些膜系统。在染色质外面生出一个核膜，将染色质包装起来成为细胞核，在核外面为细胞质，它是蛋白质合成的场所，以 RNA 为模板在核糖体(ribosome)上合成蛋白质。

最后，原始的真核细胞是不能进行有氧代谢或光合作用的。它们与好氧细菌或光合细菌形成共生关系(symbiotic association)，有些好氧细菌变成现代的真核生物中的线粒体，进行有氧代谢；另一些光合细菌与真核细胞共生，形成它们的叶绿体，进行光合作用，利用日光能将 CO_2 固定，并释放氧气，供需氧生物所利用。

第三节　原核细胞

原核细胞的特征是：它的遗传系统比较简单，没有形成由核膜包围的细胞核，只有一个环状的脱氧核糖核酸分子(circular DNA)，不含组蛋白，没有核仁，称为类核体(nucleoid)，它不进行有丝分裂(mitosis)，只进行二分裂(binary fission)，是一种无性生殖过程。

原核细胞的细胞质也比较简单，其中不存在线粒体、叶绿体、内质网、高尔基体、液泡等。但是，在原核细胞细胞质中分布有大量核糖体(ribosome)，而在真核细胞中核糖体则排列在内质网上。此外，原核细胞的核糖体体积较小，其沉降系数为 70 S，而真核生物为 80 S，都是由大亚基(large subunit)和小亚基(small subunit)组成的。原核细胞核糖体的大亚基为 50 S，小亚基为 30 S，而真核细胞核糖体的大亚基为 60 S，小亚基为 40 S。

在细胞质的性质方面，原核细胞与真核细胞也有所不同，如原核细胞中有细胞质流动(cytoplasmic streaming)及胞饮作用(phagocytosis)，而在原核细胞中则不存在。

原核细胞中呼吸作用的电子传递链(electron transport chain)及光合细菌的光系统都存在于质膜之上，而在真核细胞中呼吸作用的电子传递链存在于线粒体膜上，光合作用的光系统存在于叶绿体的类囊体(thylakoid)膜上。

原核细胞的细胞壁是由肽葡聚糖(peptidoglycan)构成的，而真核细胞(植物及菌类)的细胞壁则由纤维素和半纤维素等组成，在动物细胞中不存在细胞壁。

原核细胞(如细菌)运动的细胞器是简单的纤丝，而真核细胞的鞭毛或纤毛是具有“9＋2”结构的微管(microtubule)所组成。

在代谢方面原核细胞与真核生物也有显然不同的区别。原核细胞的代谢途径具有多样性，特别是在无氧产能反应方面，有多条代谢途径，而在真核细胞中则以糖酵解、三羧酸循环(TCA cycle)及呼吸链为主要的产能途径。

从上面简略介绍的情况可以看出，原核生物的特征非常适合分子生物学的研究。原核生物在进化上是原始的生物，它的遗传物质比较简单，基因组很小，便于操作，因此是一种极方便的研究模型。原核生物特别是细菌生长繁殖极其迅速，便于培养，作为研究材料是真核生物无法比拟的。细菌中的核酸、蛋白质及酶类都比较容易分离纯化，用它研究基因的分子生物学十分方便。近年来，分子生物学的迅速发展也是与原核细胞作为研究材料分不开的。为此，我们在本节中较详细地讨论细菌特别是大肠杆菌的一些生物学特征，以便更好地学习以后各章的内容。

(一)细菌的形态与结构

细菌是极小的微生物，大多数细菌的直径为 0.1～5.0 μm，长 2～8 μm。由于它们的细胞

极小，其表面积/体积之比(surface/volume)则特别大，这是与真核细胞有显著区别的特点。个体小有许多生理上的优点，营养物及废弃物进出细胞的速率与细胞大小成反比，从而物质运转速率影响其代谢速率及生长速率，因此，细胞越小，其潜在生长速率越快。

细菌由于为一层坚硬的细胞壁所包围，乃赋予它们以特有的形状。有些细菌呈球形或卵形，称为球菌(*Coccus cocci*)；有些细菌呈杆状，称为杆菌(*Bacillus bacilli*)；还有一些细菌呈螺旋状，称为螺菌(*Spirillum spirilla*)，见图 3-1。如果用溶菌酶(lysozyme)处理细菌，胞壁组分被去掉，胞壁的刚性即失去，这时一切细胞都会变成圆球形，最后即破裂。

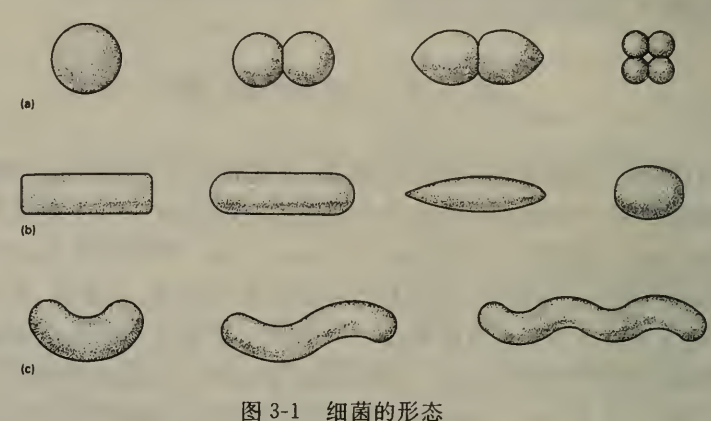


图 3-1　细菌的形态

(a)球菌；(b)杆菌；(c)螺菌。

(二)细菌的结构

细菌主要由细胞壁、细胞膜、核糖体、类核体等组成。但是，通常在细胞壁以外还含有胞壁外的结构，包括细胞糖萼、鞭毛、菌毛等。现在，我们依次予以介绍。

1. 细胞糖萼　细胞糖萼(glycocalyx)是包围细菌细胞物质的总称。细胞糖萼为胶质高分子，由多糖、多肽或二者混合组成。不同种细菌的细胞糖萼的化学组成变异很大，细胞糖萼的粘性很大，它在细胞的内部形成，而分泌到细胞表面。附着有粘性的细胞糖萼的细菌能使细菌固着在宿主组织上，这种粘性物质有时也称为荚膜(capsule)。有些细菌的荚膜对细菌致病性(virulence)很重要。荚膜可以保护这些致病菌免受宿主细胞的吞噬作用(phacocytosis)，吞噬作用是白细胞吞食并破坏细菌的一种过程。粘的细胞糖萼的另一个功能是将细菌附着在各种表面上，以便在自然环境中存在。通过附着细菌可以在物体的表面上生长，如流水中的岩石、植物根部、动物或人的组织等。

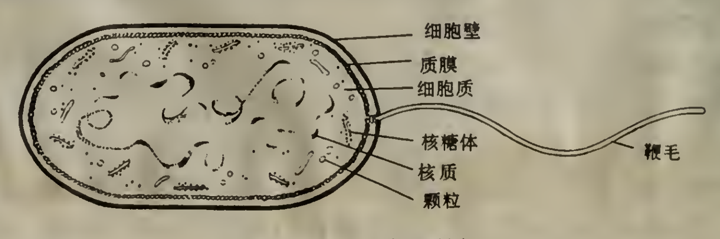


图 3-2　*E. coli* 细胞结构

2. 鞭毛　细菌的鞭毛(flagellum)为毛状，是从细胞伸出的螺旋状附属物，它担负细胞运动的功能，细菌的鞭毛比真核生物的鞭毛或纤毛(cillum)纤细得多，直径只有 0.01～0.02 μm，其

结构也较为简单(如图 3-2),细菌的鞭毛在光学显微镜下一般看不到,只有在负染后在电子显微镜下才可以观察到。

鞭毛的位置随细胞种类而变化,有的为极生(polar),在细菌的一端或两端生有鞭毛;有的为侧生(lateral),鞭毛生于两侧。鞭毛由三部分组成:①基部(basal body),它与细胞膜及细胞壁相结合;②一个短的钩状物(hook);③一螺旋状的丝(filament)。鞭毛的长度通常为细胞的数倍,基部的化学成分尚不很清楚,而钩状物及丝是由蛋白质亚基组装成的,排列成螺旋状,组成鞭毛丝的蛋白质,称为鞭毛蛋白(flagellin)。

带有鞭毛的细菌能够运动,鞭毛的运动是靠鞭毛在其基部旋转(rotation),从而推动细菌的运动,鞭毛旋转的方向为顺时针方向或反时针方向(图 3-3)。

与真核生物的鞭毛或纤毛不同,细菌的鞭毛蛋白分子没有 ATP 酶活性。因此,在细菌鞭毛做旋转运动时,ATP 并非其直接的能源,鞭毛运动的动力(或马达 motor)是由质子动势(proton motive force)驱动的,即靠跨过质膜的质子梯度所获得的动力驱动的。

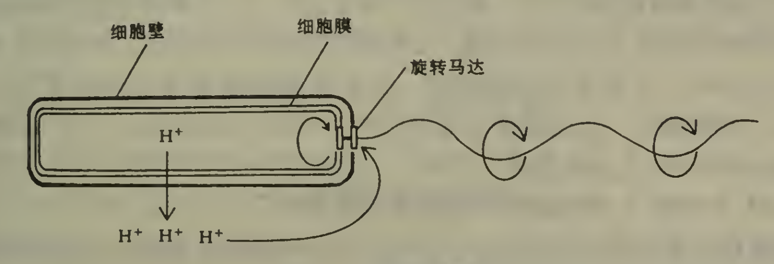


图 3-3 细菌的鞭毛

3. 菌毛 菌毛(pili)是比鞭毛更细更短而数目多的附属物,在电子显微镜下可以观察到。菌毛是非螺旋状的丝状物,没有运动的功能,因为不论在能运动和不能运动的菌种,菌毛都存在。有一种性菌毛(F-pili)在细菌交配时起作用,能使遗传物质(质粒)通过菌毛传送到另一细胞中。还有些菌毛则在人体感染时起作用,它们使致病菌能接触在呼吸道、肠道或生殖道上,这种接触能使细菌不致被体液冲洗掉,从而完成感染过程。

4. 细胞壁 在细菌外结构如荚膜、鞭毛、菌毛之下及细胞膜之外是一层细胞壁(cell wall),它是一种很坚硬的结构,使细胞保持一定的形状。它的主要功能是防止细胞吸水膨胀,以致最终破裂。因为大多数细菌生活在低渗环境中,由于吸水可能胀破。

在真细菌(eubacteria)细胞壁里,决定形状的物质主要为肽葡聚糖(peptidoglycan),它是一种不溶于水的多孔的交联高分子,有很大的刚性。肽葡聚糖只在原核生物中存在。肽葡聚糖的结构和组分在种间虽有不同,但是它们的基本结构相似,都含有 N-乙酰葡萄糖胺、N-乙酰胞壁酸、*L*-丙氨酸、*D*-丙氨酸、*D*-谷氨酸及二氨基酸(如二氨基庚二酸 diaminopimelic acid)。

通常细菌分为两大类,即革兰氏阳性(Gram-positive)及革兰氏阴性(Gram-negative),这是根据一种特殊染色法,将细菌涂片固定后加革兰氏染料染色(染料含有亚甲蓝、结晶紫 crystal violet 及番红 safranin),革兰氏阳性细菌呈紫色,革兰氏阴性细菌呈红色。细菌胞壁的结构与革兰氏染色反应有密切关系。在革兰氏染色反应中,在细胞内部形成一种不溶解的结晶紫-碘复合物,这种复合物可以用乙醇从革兰氏阴性菌提取出来,但是不能从革兰氏阳性菌提取出来。革兰氏阳性细菌的胞壁很厚,由多层肽葡聚糖组成,乙醇使之脱水,这使得胞壁上的孔关闭,阻止不溶解的结晶紫-碘复合物从细胞出来,因此染成紫色。而在革兰氏阴性细菌中,溶剂

很容易溶解出来，透过外层，薄的肽葡聚糖层也不阻止溶剂的透过，经过番红复染后，革兰氏阴性细菌乃染成红色。因此，革兰氏染色反应不是由于细胞壁的化学组成而是由于其物理结构所致。

5. 细胞膜　紧连在细胞壁下面的是一层细胞质膜(cytoplasmic membrane)，这个结构大约厚 7.5 nm，主要是由磷脂(占 20%～30%)及蛋白质(占 60%～70%)组成的，磷脂形成一个脂双层，而大部分蛋白质则镶嵌于其中。这些蛋白质称为整体蛋白(integral protein)，与脂双层结合在一起，只有用去污剂处理后将磷脂去掉时才能提取得到，其余蛋白质属于外围蛋白(peripheral proteins)，用温和处理，如渗透冲击(osmotic shock)即可除掉。膜的脂质部分有流动性，能使其组分侧向移动。流动性对膜的功能很重要，它随温度及磷脂中不饱和脂肪酸对饱和脂肪酸的比值而改变。

细菌细胞膜的主要功能有二：①ATP 的合成。细菌细胞中没有线粒体和叶绿体，根据化学渗透学说(Mitchell，1961)，在细菌中质子通过细菌的细胞膜，通过载体分子的电子流，可以驱动带正电荷的 H^+。质子运动的结果，使介质的酸度增加，产生 pH 梯度，H 离子运动导致电位差的形成。电位差是靠跨过膜泵出 H^+离子而产生的(图 3-4)；②物质的跨膜运输。通过细胞膜的物质都不是被动的运输，而是主动运输。细胞膜的透性是选择性的。细胞膜含有专一的载体蛋白(carrier protein)，它们参与运输过程。大多数膜的运输蛋白是跨膜蛋白(transmembrane proteins)，跨膜蛋白一部分暴露在细胞质外，一部分在细胞质内。由于运转蛋白的构象变化，使与细胞外表面结合的溶质进入细胞。大多数运转过程要消耗 ATP，以便克服浓度梯度。葡萄糖、果糖、甘露糖、β-葡萄糖苷等的运转是靠基团移位(group translocation)，当运转时这些糖先被磷酸转移酶系统(phosphotransferase system)使之磷酸化，然后磷酸化的糖类运入细胞，在那里它可以被正常的降解途径所代谢。

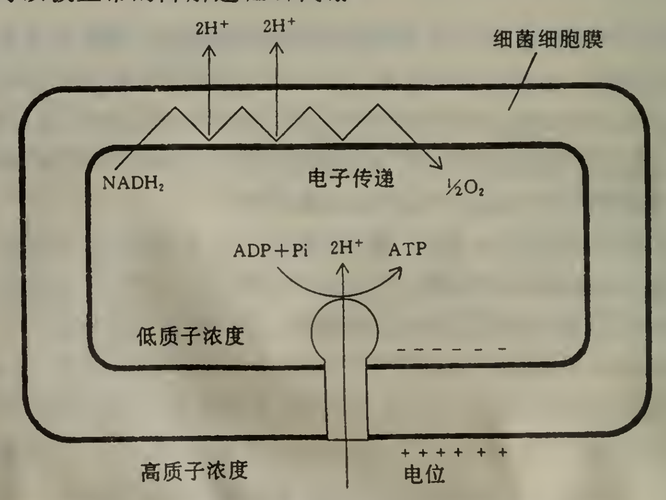


图 3-4　细菌细胞膜电位差的产生和 ATP 的生成

6. 间体　细菌的质膜常常含有一个或多个大的不规则的折叠，称为间体(mesosome)。通常间体位于或靠近细胞分裂的部位，间体在真核细胞中是不存在的。

间体主要在细菌繁殖中起作用。当细菌细胞分裂时(二分裂，binary fission)，靠间体形成一个横隔(transverse septum)，将亲代细胞的遗传物质分配到两个相同的子细胞之中，间体可能是形成横隔的开始，并使细菌 DNA 与质膜接触，间体帮助把 DNA 分配到两个子细胞中。

7. 细胞质　在细菌中细胞质(cytoplasm)被细胞质膜包围起来。它可以分为三部分：①细

胞质区，外表为颗粒状，富含大分子的 RNA 蛋白质体称为核糖体(ribosome)，它们是合成蛋白质的重要场所。原核细胞核糖体的沉降系数为 70 S，由两个亚基组成，大亚基为 50 S，小亚基为 30 S；②染色质区，其中富含 DNA。细菌与真核细胞不同，它没有明显的由核膜所包围的细胞核，也没有有丝分裂器(mitosis apparatus)，而是在细胞核的中央集中着 DNA 分子。因为它不是一个完整的细胞核，故称类核(nucleoid)或染色质体(chromatin body)或细菌染色体(bacterial chromosome)，它是由一条环状 DNA 分子组成的，所有的基因在其上连成一串。原核细胞的染色质体与真核细胞不同，它不含有组蛋白(histone)；③液体部分，在细胞质的液体部分中含有溶解的物质，如糖类、氨基酸等。与动物或植物细胞不同，细菌细胞没有核糖体结合在上面的内质网。有些核糖体游离在细胞质中，有些则与细胞质膜的内表面结合，它们参与蛋白质的合成，并将这些蛋白质运送到细胞外面。细菌的细胞质中含有成千上万的核糖体，它们使细胞质具有颗粒状的外观。

此外，细胞质的液体部分中，还含有多糖颗粒、脂质、异染粒(carboxsome)、硫颗粒等，这些统称为内含物(inclusions)。

第四节 大 肠 杆 菌

一、大肠杆菌的重要性

在原核生物中我们了解得最清楚的是大肠杆菌，我们对于 *E. coli* 的生物学的了解比任何生物都多，甚至包括人类(*Homo sapiens*)自己。*E. coli* 的结构与功能常常被当作是所有生物的原型(archetype)，*E. coli* 之所以如此出名，主要是因为它在实验室中容易操作，甚至那些对灭菌技术及其微生物学操作不熟练的人通常也能用 *E. coli* 工作而不感到困难。*E. coli* 生长迅速，只要求简单的营养物质，并且能进行很多生理生化过程，还有一点使它在现代生物学中有用，那就是性别(sex)。*E. coli* 是发现存在有性生殖的第一种细菌，1946 年 Lederberg 及 Tatum 首次发现。由于有性生殖使得遗传学的研究成为可能，遗传杂交得以进行，并且遗状性状可以分析。*E. coli* 另一有价值的性质是它能够供应细菌病毒的生长，这就使病毒的本性及病毒扩增的深入研究成为可能。因此，由于其适于实验室操作，有利于研究病毒学及有性生殖，分子生物学家及生物化学家乃能深入探讨生命的本质，他们的研究工作最终导致今日的分子生物学。

E. coli 是 1885 年由德国细菌学家 Theodor Escherich 首次发现的，它是肠道中普通的栖息者。Escherich 将这种细菌命名为大肠杆菌(*Bacterium coli*)，表明细菌为杆状，并且在大肠中栖息。以后将其属名 *Bacterium* 改变成 *Escherichia*，以纪念发现者。虽然大多数 *E. coli* 菌株是不致病的，但是有些 *E. coli* 则致病，引起腹泻及尿道感染，致病的 *E. coli* 菌株与不致病菌株有明显差异。

本节我们主要讨论与分子生物学有密切关系的一些 *E. coli* 的特点。

二、大肠杆菌的菌株

E. coli 为杆状菌，其大小为 0.5×2 μm，属于肠细菌科(Enterobacteriaceae)，*Escherichia* 属。它是兼性厌气菌，革兰氏阴性，用番红染色显红色。

E. coli 有许多菌株(strain),其野生型为 C-1A。*E. coli* 很容易培养,它的培养基成分比较简单,表 3-3 为 *E. coli* 培养基的组成。

表 3-3 *E. coli* 培养基的成分

K_2HPO_4	7.0 g	葡萄糖	10 g
KH_2PO_4	2.0 g	微量元素	每种 2～10 μg
$(NH_4)_2SO_4$	1.0 g	(含 Fe,Co,Mn,Zn,Cu,Ni,Mo)	
$MgSO_4$	0.1 g	蒸馏水	1 000 ml
$CaCl_2$	0.02 g		

三、大肠杆菌的遗传物质

大肠杆菌的遗传物质主要是染色体DNA,另外,还有一些质粒DNA,下面我们分别加以介绍。

1. *E. coli* 基因图　*E. coli* 的 DNA 为双股环状 DNA,含有 1 400 个以上的基因,这些基因都已经定位,图 3-5 就是 *E. coli* 的基因图(图中仅列出少数基因)。

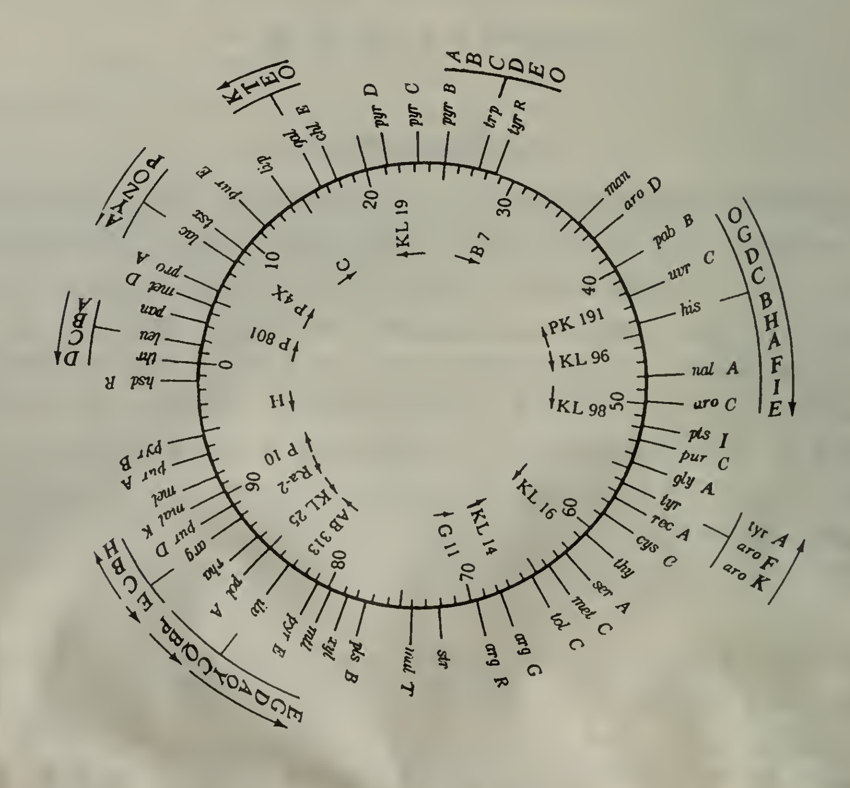


图 3-5　*E. coli* 染色体的基因图

在基因图中有些基因是集中成簇的,如操纵子基因(见第十章),但在某些相关的基因中,如组氨酸生物合成所包括的基因紧紧连在一起,但嘌呤生物合成所包含的基因则沿染色体分散(*pur* 基因)。

E. coli 的基因组(genome)的总长度为 4.6×10^6 bp,DNA 的总长度为 1 100～1 400 μm。

2. 质粒　质粒(plasmid)是细菌中的另一类遗传物质。质粒是环状 DNA,能自我复制,它

们存在于染色体之外。质粒能靠结合过程(conjugation process)从细胞代代相传，某些质粒还能与染色体整合(integration)，在这种情况下其复制则受染色体的控制。

质粒也携带有许多基因，例如，控制产生毒素的基因，对抗菌素、重金属及其它抑制物有抗性的基因。有些质粒携带着降解稀有底物如芳香化合物及杀虫剂的基因。图 3-6 为 pBR322 的基因图。

pBR322 是 *E. coli* 中的质粒，长 4 367 核苷酸，是分子生物学研究中最常用的一种质粒，其上有各种内切酶如 *Bam*H I, *Pst* I, *Sal* I, *Eco*R I, *Hind* Ⅲ 等的酶切位点。此外，*E. coli* 的质粒的种类很多，如：

pAT 153 (36 kb)，　　pKC 7 (58 kb)，
pMK 16 (46 kb)，　　pCR 1 (11.4 kb)。

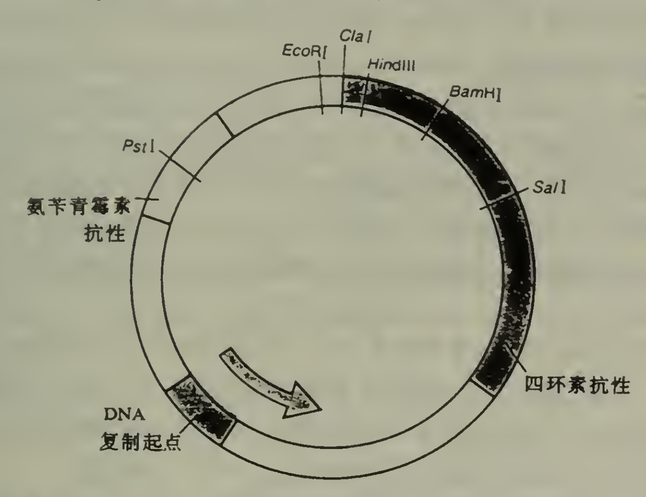


图 3-6　质粒 pBR322 的基因图
箭头表示 DNA 从原点复制的方向。

四、大肠杆菌的酶类

E. coli 细胞中含有多种酶类，除去进行合成代谢及降解代谢所需要的酶类之外，*E. coli* 还含有核酸代谢所需要的酶类，包括 DNA 聚合酶(DNA polymerase)、RNA 聚合酶(RNA polymerase)以及多种限制性内切酶(restriction endonuclease)。限制性酶只在具有特异顺序碱基的位点上与 DNA 结合。限制性酶有一个特点，它能在双链上沿一固定位点按对称顺序打断 DNA。例如，*E. coli* 的一种限制性内切酶 *Eco*R Ⅰ，具有下列识别顺序：

5′……G—A—A—T—T—C……3′
3′……C—T—T—A—A—G……5′

裂解的部位以箭头表示，对称轴以虚线(……)代表。

限制性酶对 DNA 的研究极其重要，因为它们可使大的 DNA 分子切割成小的片段，由限制性酶所生成的片段可以测定其序列，从而可以完成 DNA 分子的全序列测定(详见第四章)。现在，许多种限制性酶已有商品出售，其中不少是从 *E. coli* 细胞分离纯化得到的，成为研究分

子生物学必不可缺少的工具酶。

从 *E. coli* 细胞分离得到的限制性内切酶常用的还有下列各种：*Eco*R Ⅰ，*Bam*H Ⅰ，*Hind*Ⅲ，*Pst* Ⅰ，*Sal* Ⅰ，*Sma* Ⅰ 等等。上述各种聚合酶及限制性内切酶将在以后各章中经常遇到。

第五节　真核细胞

地球上的生物大多数属于真核生物，包括原生动物(protist)、真菌(fungi)、动物和植物。从单细胞的酵母(*Saccharomyces*)到人类都是真核生物。我们在本章的第二、三节中已经谈到过真核生物区别于原核生物的一些特征。真核细胞中存在着明显的细胞核，外面有一层核膜，其中的脱氧核糖核酸(DNA)与组蛋白结合形成核小体(nucleosome)。这些遗传物质在有丝分裂或减数分裂时浓缩成特定数目的染色体。在真核细胞的细胞质中还存在着许多种细胞器，如线粒体、叶绿体、内质网、高尔基体、溶酶体等，承担着呼吸作用、光合作用、蛋白质合成等重要功能。真核细胞中还有一个由微丝、微管及中间纤维组成的细胞骨架系统，在细胞质中构成一个复杂的网络，维持不同细胞所特有的形状，并执行细胞运动和物质运输等功能。

酵母是单细胞的真核生物。酵母细胞虽然很小，但是它具有真核细胞的结构和功能。酵母的遗传物质 DNA 分布在 16 条染色体上，共含有 11×10^6 bp，它能进行有丝分裂和减数分裂。酵母的细胞质中含有各种细胞器，包括线粒体、粗糙内质网、高尔基体、溶酶体和细胞骨架(微丝和微管)。

酵母细胞中除存在 DNA 外，还含有各种 RNA(包括 rRNA，tRNA 和 mRNA)及 DNA 聚合酶、RNA 聚合酶等，它能进行 DNA 的复制、转录及蛋白质合成。因此，酵母也是研究分子生物学的一个很重要的材料，并且它有易于培养、容易操作、便于大量繁殖等优点，通常用作基因表达的载体，其表达产物(蛋白质)可以分泌到细胞之外，便于分离纯化。

酵母中富含线粒体，因此，酶母曾经是研究糖酵解(glycolysis)等生物氧化作用的重要材料，对生物化学研究做出了重大贡献。

衣藻(*Chlamydomonas*)也是一种单细胞的真核生物，它属于眼虫目(Euglenoid)。衣藻细胞上有两条鞭毛，能在水中游泳。其 DNA 为环状。它的细胞质中除含有线粒体外，还含有叶绿体，能进行光合作用。它的叶绿体中含有光合作用所需的各种成分，如叶绿素、质体蓝素、光系统Ⅰ、光系统Ⅱ、P700 等。衣藻在实验室中很容易培养，因此，衣藻成为研究分子生物学常用的良好材料。

在真菌的粘菌科(Myxomycete)中有两种粘菌(slime mold)最为常见，一种是单细胞粘菌(cellular slime mold)，即盘基网柄菌(*Dictyostelium discoideum*)，它的细胞呈变形虫状，当细胞发育到一定时期，由于环式 AMP 的作用，细胞集合起来，形成一个能运动的聚合体，以后产生子实体，生出孢子，孢子萌发即是变形虫状的细胞；另一种是多头绒泡菌(*Physarum polycephalum*)，它的细胞集合成一个多核的原质团(plasmodium)。当生存条件不良(如干燥)时，它会产生孢子。这两种粘菌是研究发育的分子生物学的良好材料。

在多细胞真核生物中许多种动物和植物现在都成为分子生物学的研究材料，在重组 DNA 技术方面已不成为问题。不过，在遗传学和发育生物学方面果蝇(*Drosophila melanogaster*)的研究是特别深入的。在果蝇的变态过程中，从受精卵到幼虫、蛹，以至于孵化成成虫，受着很多种基因在时间、空间和序列的调控，是一个非常复杂而奥妙的问题，现在已经有不少过程研究

清楚了。

在植物方面，拟南芥(*Arabidopsis thaliana*)是一个研究分子生物学的理想材料。拟南芥属于十字花科(Crucifer)，它的染色体 $2n=10$。由于很多科学家的努力，拟南芥的基因组序列已经测定出 150 000 bp 以上，并存入基因银行中，可供查阅。

下面我们较详细地介绍真核细胞的细胞核、细胞质、质膜等的结构。

一、细　胞　核

真核细胞的细胞核(nucleus)比原核细胞的类核体(nucleoid)在结构上和功能上都复杂得多。细胞核含有大部分的 DNA，只有一小部分 DNA 存在于线粒体或叶绿体中。细胞核被核膜(nuclear envelope)所包围。核膜由两层膜组成，其外层与内质网(endoplasmic reticulum)相连接。

在细胞核内存在着 1 个或多个核仁(nucleolus)。核仁是 DNA 转录成核糖体 RNA 的场所。RNA 合成后通过核孔运送到细胞质中，再组装成核糖体。除核仁部分外细胞核的其余部分为染色质，它们由 DNA 及与之紧密结合的蛋白质组成，在有丝分裂(mitosis)的前期，它们浓缩成染色体(chromosome)(图 3-7)。各种动植物染色体的数目和大小不同，随物种而异，如人类的染色体数为 46，四膜虫(*Tetrahymena*)为 4，玉米(*Zea mays*)为 20，拟南芥为 10。与 DNA 结合的蛋白质主要是组蛋白(histone)，它是一种碱性蛋白质，能与呈酸性的 DNA 大分子相结合。此外，细胞核中有时还存在酸性的非组蛋白(nonhistone)。组蛋白与 DNA 结合成一种复合体称为核小体。

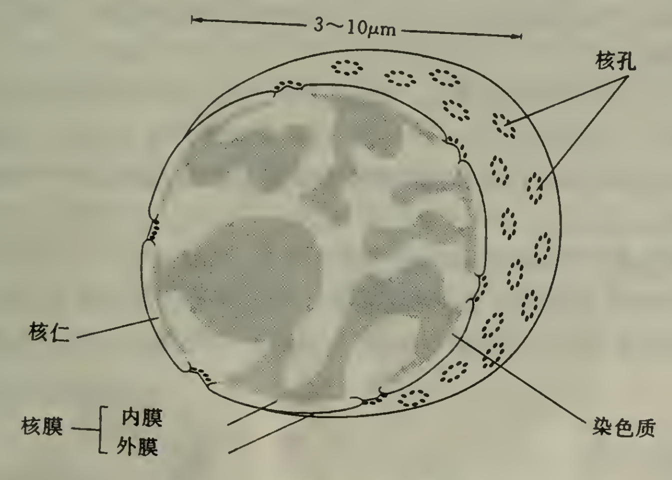


图 3-7　细胞核的结构示意图

在原核细胞和真核细胞 DNA 大分子中含有大量基因，原核生物如支原体(*Mycoplasma genitalium*)的 DNA 为一环状分子，所有基因都排列在上面，构成它的基因组(genome)。在真核细胞中基因组很大，于是形成许多个染色体，每个染色体为一个 DNA 分子，共同构成整个基因组。基因组含有编码各种蛋白质、酶、RNA 的基因(DNA)，这些基因控制着整个细胞的结构成分与功能，如碳水化合物、脂肪，氮素代谢及能量代谢的途径，DNA 的复制、转录、翻译、调控，以及细胞的其它功能。

基因组是当前生物学研究的重大课题，特别是“人类基因组计划”(Human Genome Pro-

ject)受到国际上普遍重视，世界上许多国家都投入了大量人力、物力、财力，正在采用最先进技术紧张地进行研究。由于人类基因组大约含有10^5个基因，3×10^9个bp，全部基因组的测定需要很长的时间。基因组控制着人类的全部遗传信息和各种生命活动，一旦基因组的基因定位、核苷酸序列及氨基酸序列都测定完成，我们对这些基因的功能将会更好地理解，并且在医学上对遗传病治疗、基因治疗等将做出极其重大的贡献。

目前，许多科学家已经完成一些微生物、动物、植物或细胞器基因组的核苷酸全序列，我们将这些结果列于表3-4。

表3-4　一些微生物、动物和植物基因组的研究结果

生　　物	碱基对数(bp)	作　　者
嗜菌体 *ΦX*174 (*Bacteriophage ΦX*174)	5 386	Sanger (1977)
嗜菌体 λ (*Bacteriophage* λ)	48 502	Sanger (1982)
CMV (*Cytomegalo virus*)	229 000	
Mycoplasma genitalium	580 070	Venter 等 (1995)
Hemophilus influengae	18 301 376	
线虫 *Caenorhabditis elegans*	100 000 000	Hodgkin 等 (1995)
拟南芥 *Arabidopsis thaliana* 染色体 4	15 000 000	
烟草叶绿体 *Nicotiana tabacum*	155 844	杉浦昌弘等 (1986)

在Sanger测定了嗜菌体*ΦX*174，*λ*的基因组之后，近年来Venter等(1995)完成了支原体(*Mycoplasma genitalium*)基因组的全序列分析。*Mycoplasma*的基因组为一环状DNA，据认为是细菌中最小的基因组，共有580 070 bp，其中G+C占32%，A+T占68%，此项研究工作量很大，由30人完成，将基因组切成8 472个片段分析后再组装得到。他们认为所编码的蛋白质和酶都已鉴定出来。

在动物方面线虫的研究最为深入。线虫是研究发育生物学的模式动物。它的个体很小，约1mm长，身体透明，生活史短，3天一个世代，可在琼脂培养基上生长。它的遗传物质DNA约为1亿bp，为人类基因组的1/30。它的物理图谱较易构建，截至1995年为止，线虫DNA全序列的1/5已经测定完成。

在植物方面拟南芥基因组的研究是当前的重点。拟南芥共有10条染色体。现在，重点研究第4条染色体DNA的序列。在我国水稻基因组的研究已列为重点，现正进行研究。当水稻基因组的全部基因及DNA序列测出后，我们对水稻的遗传性及基因工程的研究将会有更大的改进。

二、细　胞　质

在一个细胞的质膜之内，除去细胞核以外，都属于细胞质(cytoplasm)的范畴。细胞质中含有多种细胞器(organelle)，最显著的是线粒体，它是供应细胞能源的中心。细胞的各种生命活动如生物合成、运动、物质运输所需要的能量都是靠线粒体将碳水化合物等通过糖酵解、三羧酸循环及电子传递链以及氧化磷酸化等步骤形成腺苷三磷酸(ATP或GTP)。ATP是一切生物可利用的能的形式。在绿色植物中还存在一种叶绿体，它是植物进行光合作用的场所。植物叶片依靠叶绿体中的1,5-二磷酸核酮糖羧化酶(Rubis CO)将CO_2固定，并通过卡尔文(Calvin)循环生成糖类，同时放出氧气，这是植物所需氧气的重要来源。

细胞质中还存在进行蛋白质合成的内质网(endoplasmic reticulum)、进行蛋白质加工的

高尔基体(Golgi body)、溶酶体(lysosome)、过氧化物体(peroxisome),在植物细胞中还有液泡(vacuole)。在细胞质中还有一个细胞骨架系统,它由微丝、微管和中间纤维组成一个网络系统,对维持细胞形状、进行运动有重要功能。除去这些结构以外就是一些溶液,称为胞液(cytosol),其中含有各种有机物质和无机离子。下面我们扼要地讨论主要的细胞器。

(一)线粒体

在所有真核生物的细胞中都含有线粒体(mitochondria),其形状为线状或粒状。典型的线粒体的大小约为 1 μm,与细菌细胞相似,大多数动植物细胞含有数百个到 1 000 个线粒体。线粒体通过其中的生物氧化作用将糖类分子中的化学能转变成ATP,供细胞维持生理功能之用,因此,线粒体是细胞的发电厂。

每个线粒体有两层膜,外层包围整个线粒体,内层向内折叠成内嵴(crista),赋予它以大的表面积,线粒体的嵴与嵴之间为间质(matrix),其间充满很浓的酶类,如三羧酸循环的酶类及中间代谢物。进行电子传递的蛋白质及氧化磷酸化的腺苷三磷酸酶(ATP酶)分子则镶嵌在内膜之中。ATP 酶分子的三维结构现在已经用晶体衍射技术解析出来(图 3-8)。

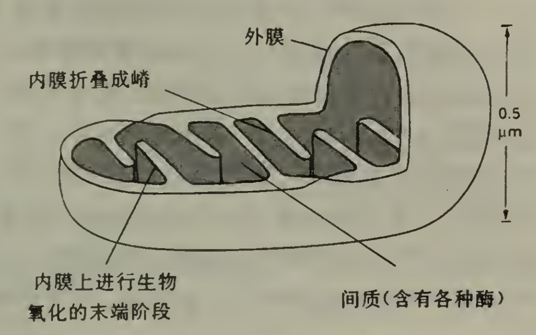


图 3-8　线粒体的结构示意图

线粒体有其自己的基因组,它编码构成线粒体的蛋白质、酶类和 RNA,但是也有些线粒体蛋白质由核基因编码。

由于线粒体有其自己的基因组,这也有助于证明线粒体是由远古的好氧细菌与真核生物共生的理论。

(二)叶绿体

在绿色植物细胞中含有许多质体(plastids),其中最显著的是叶绿体(chloroplast)。与线粒体相似,叶绿体是绿色细胞的发电厂,它将叶绿素吸收的日光能转化成ATP分子。叶绿体的叶绿素吸收日光能通过光系统反应中心激发出电子,经过电子传递链,使 ADP 磷酸化成为 ATP。叶绿体中含有大量1,5-二磷酸核酮糖羧化酶/加氧酶,将 CO_2 固定为己糖,再通过卡尔文循环生成葡萄糖,以后葡萄糖可以转化成蔗糖和淀粉。光合细菌的反应中心的晶体结构已于 1989 年由德国的 Deisenhofer,Michel 和 Huber 解析出来,获得了 Nobel 奖。图 3-9 为叶绿体的示意图。

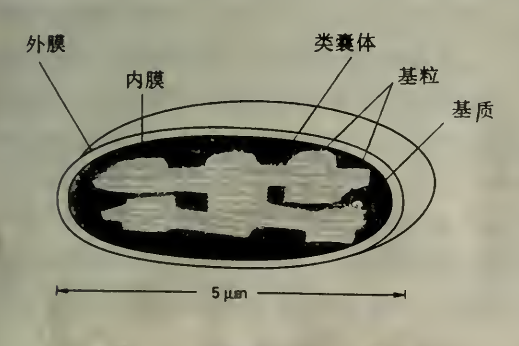
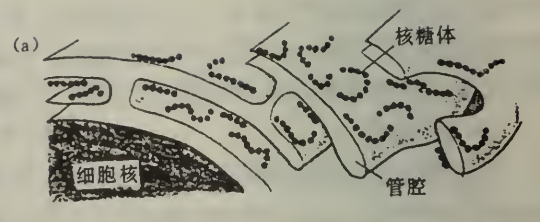


图 3-9　叶绿体的结构示意图

日本的杉浦昌弘等人于 1986 年将烟草(*Nicotiana tabacum*)叶绿体基因组 DNA 的全序列测定成功,它由 155 844 bp 组成。这些核苷酸序列组成了编码叶绿体中所有蛋白质(如 psaA,B,psbA,B,C,D)的基因及 tRNA 和 rRNA 的模板。这项研究结果对我们认识光合作用的分子基础是至关重要的。

（三）内质网

内质网是细胞中合成蛋白质的重要场所。内质网上有很多合成蛋白质的核糖体(ribosome)。内质网是贯穿整个细胞质的由膜组成的网络，与胞质分开.其表面有许多核糖体，显得很粗糙，因此，常称为粗糙内质网(rough endoplasmic reticulum)。它有许多扁平状的分枝，彼此联系并与核膜结合(图 3-10)。在特化成专门分泌蛋白质的细胞(如胰脏分泌胰岛素的细胞)中，内质网特别丰富。合成脂质的内质网表面是光滑的，称为平滑内质网(smooth endoplasmic reticulum)。在骨骼肌中内质网特化成贮藏及释放 Ca^{2+} 的结构，Ca^{2+} 离子释放是肌肉收缩以及许多生理功能的扳机(trigger)。合成向外输出的蛋白质的内质网定位于内质网的外表面(朝向胞质)，而在那些存留在胞液(cytosol)中的蛋白质则是在与内质网不结合的核糖体上合成的。有关核糖体的结构与功能还将在第九章中详细讨论。



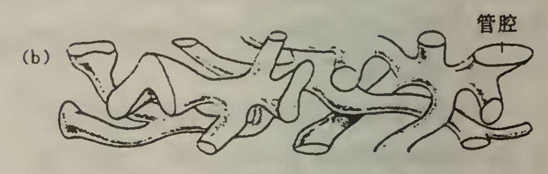


图 3-10　内质网的结构示意图

(a)粗糙内质网(含核糖体)；(b)光滑内质网。

（四）高尔基体

真核生物细胞有一个由膜囊泡组成的结构称为高尔基体(Golgi body or complex)。一个高尔基体通常可看作是一个扁平的膜囊泡(内嵴)。靠近这些内嵴的末端有无数更小的球状囊泡，它们是运输囊泡(图 3-11)。

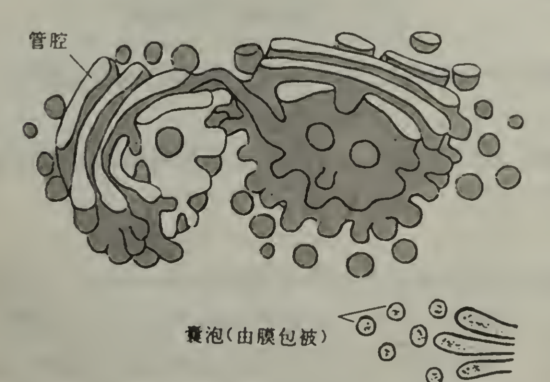


图 3-11　高尔基体的示意图

高尔基复合体在结构上和功能上都是不对称的，其顺向(cis)一侧面向粗糙内质网，而反向(trans)侧则面向质膜，二者之间为中间介质(medial element)。当蛋白质在核糖体上合成时，蛋白质嵌入到内嵴的内部。含有新合成的蛋白质小膜囊泡从内质网上以出芽方式运动到高尔基体上，与顺向侧融合。当蛋白质通过高尔基体到达顺向侧时，高尔基体中的酶类即将蛋白质分子加以修饰，如将硫酸基团、碳水化合物或脂质加到某些氨基酸侧链上。这些新合成蛋白质修饰的功能之一就是将蛋白质定位于正确的位置上，当它在运输囊泡中离开高尔基体即从反向(trans)侧生出芽。有些囊泡包容在分泌囊泡中，最终以外吐(exocytosis)方式从细胞中排出去。另一些则镶嵌在细胞内的细胞器上，如溶酶体，或者在细胞生长时掺入到质膜之中。

（五）溶酶体

溶酶体(lysosome)存在于动物细胞中，呈球状，由单层膜组成，直径通常为 1 μm。溶酶体中含有能够溶解的蛋白质、多糖、核酸及脂质的水解酶类(图 3-12)。这些生物大分子被酶类降解成简单的成分，如氨基酸、单糖、脂肪酸等，排入泡液之中，重新循环利用。

过氧化物体：在某些细胞中存在着一种过氧化物体(peroxisome)，其中含有较大量的过氧

化氢酶(catalase),它可催化下列反应:

$$2H_2O_2 \longrightarrow 2H_2O+O_2$$

在氨基酸及脂肪降解所进行的氧化反应有时产生自由基(free radical)及过氧化氢,它是很活跃的化学试剂,可以破坏细胞的结构。为了免于被这些副产物所破坏,细胞将这些反应集中到小的由膜包围的过氧化物体之中。

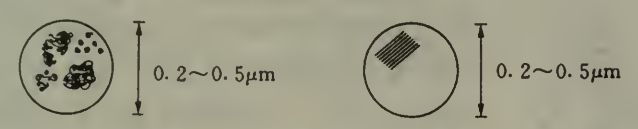


图 3-12　溶酶体及过氧化物体

(六)乙醛酸体

在油料植物细胞中存在一种特化的细胞器,它们含有很高浓度的乙醛酸循环(glyoxylate cycle)的酶类(主要是异柠檬酸裂解酶及苹果酸合成酶)。乙醛酸循环是油料植物所特有的一条代谢途径,当种子萌发时可将贮存的脂肪转化成碳水化合物。

过氧化物体、溶酶体和乙醛酸体有时总称之为微粒体(microbodies)。

(七)液泡

植物细胞不含溶酶体,它们的液泡(vacuole)能执行类似的降解反应。幼嫩植物细胞含有许多小液泡,由单层膜形成,成熟细胞中央由小液泡融合成一个大液泡。液泡膜(tonoplast)能调节离子和代谢物进入液泡(图 3-13)。在有些植物中液泡内含有高浓度的色素,如花青素(anthothyanin)、黄酮类(carotinoid)使玫瑰、葡萄等花果具有红、紫、蓝、黄等各种颜色。

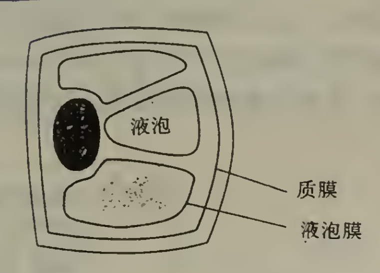


图 3-13　液泡结构的示意图

三、细 胞 骨 架

细胞质并不像过去所了解的那样是一小袋溶液(胞液),而具有非常复杂的结构。由于细胞生物学和生物化学研究方法的日臻完善,现已证明,细胞质中除含有线粒体、高尔基体、内质网等细胞器之外,还有一个组成细胞质的重要结构,即细胞骨架(cytoskeleton)。研究表明,一切真核细胞中都存在细胞骨架,细胞骨架可以使细胞形状稳定,使细胞质组织化,并且能产生运动(motility)。

细胞骨架由三部分成分组成:即微丝(microfilament)系统、微管(microtubule)系统和中间纤维(intermediate filament)系统(图 3-14)。我们分别讨论如下:

1. 微管　25 nm 直径
2. 肌动蛋白丝　8 nm 直径
3. 中间纤维　10 nm 直径

图 3-14　细胞骨架的示意图

(一)微丝系统

微丝由肌动蛋白(actin)、肌球蛋白(myosin)和肌动蛋白结合蛋白(actin-binding protein,简称 ABP)组成。肌动蛋白于 1942 年由 Straub 在肌肉细胞中发现。它的单体是一个球形分子,称为 G-actin,

分子量为 42 kD，由 375 个氨基酸组成。1990 年 Kabsch 等用 X 光衍射技术测出了它的晶体结构。它的分子由 2 个结构域(structural domain)组成，每个结构域又可分成 2 个亚结构域，中间有一个裂隙，其中可以结合 ATP。许多 G-肌动蛋白分子在 KCl 存在时可以聚合成纤维状的 F-肌动蛋白，或称为肌动蛋白丝(actin filament)，亦即微丝。微丝赋予细胞以比较坚韧的结构，常称之为应力丝(stress fiber)，从而可以使细胞具有稳定的结构。

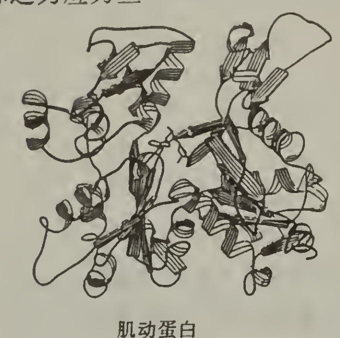


图 3-15　肌动蛋白的三维结构

肌动蛋白的基因称为 *act* 基因，目前已经有 100 多种 *act* 基因的核苷酸序列测定完成，包括脊椎动物、昆虫、原生动物、植物、藻类和真菌的 *act* 基因。这些基因是非常保守的，研究表明每 1 亿年 1 个氨基酸才能产生 1 次突变。根据这 100 多种动植物 *act* 基因核苷酸序列或肌动蛋白氨基酸序列，可以绘出生物的进化树(图 3-15)。

肌球蛋白(myosin)是肌动蛋白丝的马达蛋白(motor protein)，马达蛋白具有 ATP 酶活性，可使 ATP 分子的化学能直接转化成机械能做功，推动肌动蛋白丝运动。肌球蛋白的分子量为 460 kD，由 2 条重链和 4 条轻链组成，具有 ATP 酶活性。电子显微镜显示肌球蛋白分子有两个圆形的头部和一个长的尾部(图 3-16)。肌球蛋白的晶体结构已于 1993 年由 Rayment 等测定出来。许多生物的肌球蛋白基因(*myo* gene)如酵母、盘基网柄菌等的肌球蛋白基因的核苷酸序列也已经测定出来(图 3-17)。



图 3-16　肌球蛋白头部的三维结构

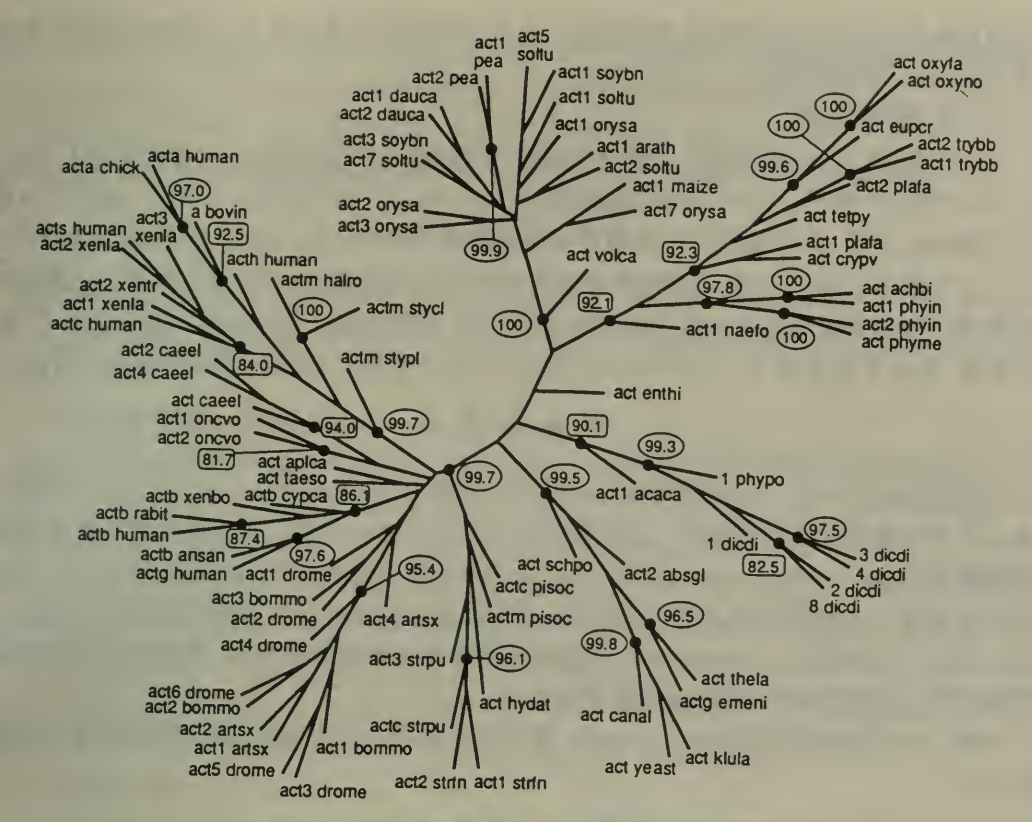


图 3-17 动物、植物、真菌中的肌动蛋白基因的进化树

动物中的肌动蛋白结合蛋白(ABP)已发现了多种,如原肌球蛋白(tropómyosin)、肌钙蛋白(troponin),肌动蛋白丝、肌球蛋白及各种 ABP 在细胞中构成一个复杂的网络系统(肌动蛋白骨架 actin cytoskeleton),维持细胞所特有的结构,并且可以进行运动,如变形虫运动、粘菌的穿梭运动、植物细胞的胞质流动等,都是由于肌动蛋白丝与肌球蛋白的相互作用引起的。

我们于 1963 年在南瓜、烟草中发现肌动球蛋白(actomyosin)的存在,并且具有 ATP 酶活性。以后,研究证明植物细胞中普遍存在肌动蛋白和肌球蛋白,与动物完全相同。

(二)微管系统

微管系统由微管(microtubule)及微管结合蛋白(microtubule associated protein,简称 MAP)组成。微管是 1964 年由 Porter 等用电子显微镜发现的。微管的直径为 25 nm,它是微管蛋白(tubulin)α 亚基与 β 亚基的聚合体,中空管状,每圈由 13 个亚基组成。微管蛋白分子量为 55 kD。在体外紫杉醇(taxol)可促进微管蛋白的聚合。在 MAP 中现已发现 3 种马达蛋白,即力蛋白(dynein)、动蛋白(kinesin)和动力蛋白(dynamin)。前二种具有 ATP 酶活性,后一种具有 GTP 酶活性。力蛋白是 Gibbons(1965 年)从海胆精子、衣藻鞭毛中发现的,它是构成鞭毛或纤毛的主要成分。它的分子量极大,约为 500 kD。近年在真核细胞的细胞质中发现了胞质力蛋白(cytoplasmic dynein)。动蛋白是 Vale(1989 年)从动物脑细胞中发现的,分子量也很大,约为 400 kD。最近,它的晶体结构也已测定出来,很像肌球蛋白,也具有两个头部和一个尾部。动力蛋白是 Vallee(1989)发现的,与力蛋白和动蛋白不同,它具有 GTP 酶活性,分子量为 100 kD。这三种马达蛋白都可以推动细胞器或蛋白颗粒沿着微管进行正向或反向的运动,在神经细胞

的长距离物质运输中起着非常重要的作用。上面介绍的微管蛋白及其马达蛋白的基因都已经克隆，并测定了核苷酸序列。

（三）中间纤维系统

中间纤维(IF)是一种直径介于微丝与微管之间的细丝状网络系统，它的直径为 10 nm。组成 IF 的单体有许多种，其中一种称为波状蛋白(vimentin)，分子量约为 57 kD。波状蛋白丝常位于细胞核的周围，其功能是将细胞核固定在细胞质的中央；另一种 IF 的单体为角蛋白(keratin)，分子量为 55 kD。它们的功能可能是将横纹肌的 Z 线(Z disc)固定在该处。在植物细胞中杨澄等(1992)发现也存在角蛋白组成的中间纤维。动物神经元(neuron)中存在三种 IF 的蛋白质亚基，其分子量分别为 70 kD，150 kD 和 210 kD，它们赋予伸长的神经元细胞以刚性。

四、细 胞 质 膜

在细胞质的外面被一层生物膜包围，这层生物膜称为质膜(plasma membrane)。质膜也是细胞与其它细胞相接触的外表面。它由脂双层(lipid bilayer)组成。在质膜上含有各种内在蛋白和跨膜蛋白(transmembrane proteins)。在电子传递链中合成 ATP 的 ATP 酶(ATPase)，就是定位在质膜上的跨膜蛋白，它在 ATP 的合成并供应细胞以能量上是非常重要的膜蛋白。在质膜上有许多运输蛋白(transporter)，它们能把营养物运输到细胞内部，并将产生的废物运送到细胞之外，对维持细胞的代谢起着重要的作用。

质膜上有许多离子通道(ion channel)，能使无机离子通过生物膜。离子通道的专一性很强(图 3-18)。

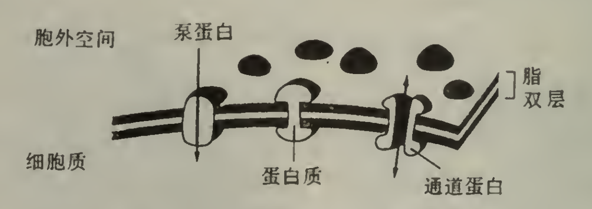


图 3-18　质膜结构的示意图

在质膜上还存在许多信号受体(signal receptor)，它能接受外界的信号分子(signal)，如光、热、触觉、味觉、嗅觉、神经递质等。当外界的刺激传导到受体上，受体即将该信号(或配体 ligand)转导到细胞内的通道，使通道开放，允许大量离子通过。在信号转导中最重要的是神经的信号转导(signal transduction)。乙酰胆碱(acetylcholine，简称 ACh)是一种重要的神经递质(neurotransmitter)，它可以将信号传递给乙酰胆碱受体(ACh receptor)。当动物受到外界的刺激如电刺激时，刺激的信号即通过神经细胞(神经元 neuron)中的 ACh 传递到 ACh 受体上，引起 Ca^{2+} 的释放，于是肌肉即发生收缩。在动物细胞的免疫系统中，当外来的细菌、病毒或蛋白质作为抗原进入细胞后，即被细胞的特殊受体识别出来，开动扳机产生抗体(antibody)。信号转导在生命过程中起着极其重要的作用。因此，目前在现代生物学中信号转导的研究成为一项热点课题。

近年来发现在细胞的质膜下面还存在一个膜骨架(membrane skeleton)结构。动物的红细胞如果在低渗溶液中胀破，血红蛋白全部流出之后，即剩下一个血影细胞(ghost)，即血红细胞空壳。在血影细胞的质膜下面发现了一些特殊蛋白质，它们是：血影蛋白(spectrin)、带 3 蛋白

(band 3 protein)、肌动蛋白(actin)、锚蛋白(ankyrin)和 4.1 蛋白。用电子显微镜及胶体金技术证明,在红细胞的质膜下面存在一膜骨架网络。这个膜骨架结构与细胞骨架有紧密的联系,在生命活动中起重要作用。现在业已证明,除红细胞以外,其它细胞,如果蝇以及植物(蚕豆表皮细胞)中也存在有膜骨架蛋白和膜骨架结构。膜骨架是真核细胞的一个重要结构。

五、细胞的基本功能

细胞是动植物和微生物的基本单位,它执行这些生物的基本功能。我们不但要了解细胞的结构,而且要了解它们的功能,才真正了解细胞的全貌。下面我们扼要地讨论细胞的一些主要的功能。

(一)细胞的分裂(cell division)

细胞分裂包括细胞核分裂和细胞质分裂两部分。在真核细胞中细胞核分裂称为有丝分裂(mitosis)。有丝分裂分为 4 个时期,即前期(prophase)、中期(metaphase)、后期(anaphase)和末期(telophase)。在前期细胞核中分散的染色质首先浓缩成染色体,每个染色体又等分为两个染色体。中期时在细胞质中形成纺锤体,染色体排列在中间的赤道板上。后期时两套染色体分别向两极的中心体移动。末期时到达两极的染色体分别形成两个子细胞的核,重新产生核膜,将染色质包装起来。纺锤丝是由微管组成的。染色体向两极的运动靠微管及其马达蛋白(动蛋白)的相互作用进行。由此可见,细胞骨架在有丝分裂及减数分裂中起着重要作用。近来还有证据表明,肌动蛋白丝和肌球蛋白在有丝分裂过程中也起着较重要的作用。

(二)胞质分裂(cytokinesis)

动物细胞在有丝分裂后在细胞的中央形成一个收缩环(contractile ring),将一个细胞分裂成两个子细胞(daughter cells),这个过程称为胞质分裂。现已证明:胞质分裂是由于细胞中央出现的肌动蛋白丝(微丝)进行收缩,将母细胞平均分割成两个子细胞。

植物细胞与动物不同,它是在有丝分裂后于细胞质的中央形成一个成膜体(phragmoplast),将两个子细胞分开。

(三)细胞运动(cell motility)

动物的精子、衣藻(*Chlamydomonas*)细胞都具有鞭毛(flagellum)、草履虫(*Paramecium*)表面具有许多纤毛(cillium),它们依靠这些鞭毛或纤毛的摆动,使这些原生动物可以在水中自由泳动。鞭毛和纤毛的运动都是靠这些细胞器中的微管系统进行的。用电子显微镜可以观察到纤毛或鞭毛的横断面上排列着一圈微管,外面是 9 个成双的微管,中间是 2 个单个的中心微管,共同形成"9+2"的特殊结构。微管与其马达即力蛋白相互作用而使纤毛或鞭毛运动。编码动植物微管蛋白和力蛋白的基因的核苷酸序列已经测定出来。

动物的肌肉收缩(muscle contraction)是一种特化的细胞运动。执行肌肉收缩的肌小节(sarcomere)细胞中含有规则排列的肌动蛋白丝(细丝)和肌球蛋白丝(粗丝)。Huxley(1969)根据他对横纹肌结构电子显微镜的大量研究,提出了一个著名的肌丝滑动模型(sliding filament model)。他认为肌球蛋白头部与肌动蛋白丝上的肌动蛋白分子接触后,利用头部的移动,推动肌动蛋白丝产生运动。后来,大量研究发现除肌肉外,动物小肠壁的刷状缘(bruch border)、白细胞、血小板(platelet)、成纤维细胞、精子等多种非肌细胞,都有类似肌肉收缩的运动和类似肌丝滑动模型的分子运动机理。例如,小肠的刷状缘细胞中排列有整齐的肌动蛋白丝,依靠这些微丝刷状缘可以不停地运动,促进小肠吸收营养物。动物精子头部的顶体(acrosome)中含有

大量肌动蛋白单体，当授精时这些单体迅速聚合成肌动蛋白丝，从而将精子的细胞核推到卵细胞中，因此，顶体中的肌动蛋白在授精作用中起着极其重要的作用。

植物细胞的运动很多也类似肌肉收缩，靠细胞中的肌动蛋白丝及肌球蛋白的相互作用，如丽藻(*Nitella*)节间细胞中的细胞质流动(cytoplasmic streaming)，各种植物花粉管中的细胞质流动，都是由于细胞中的肌动蛋白丝与肌球蛋白相互作用的结果。

最近的研究结果表明，有些细胞运动，如伸出(protrusion)只依靠肌动蛋白单体聚合成肌动蛋白丝，就可以推动细胞向前运动。

最近两年关于细胞运动的研究有了很大的突破。Spudich(1994)和柳田敏雄(Yanagida，1994)分别采用光钳(optic trap)技术研究单个肌球蛋白头部所产生的运动，取得成功。他们将肌肉的单个肌动蛋白丝的两端用光钳固定在塑料小球上，将肌球蛋白一个头部与肌动蛋白单体接触，在生理条件下，用非常灵敏的光电二极管(photodiode)装置，测量肌球蛋白推动肌动蛋白丝所产生的力和所移动的位移。他们两个研究组所得的结果基本相同。单个肌球蛋白头部可以产生 3～4 pN 的力(1 个牛顿 N 等于 10^7pN)，并使肌动蛋白丝移动 11 nm 的距离。这项研究使我们能从分子水平了解动植物的细胞运动，同时在物理学上细胞运动的研究水平使生物力学(biomechanics)取得了很大突破。

(四)物质的运输

细胞还有运输物质(如蛋白质或其它分子)的功能，特别是距离较远的时候，靠简单的扩散作用不能达到所需要的速度。例如，动物神经元(neuron)细胞结构很复杂，在它的细胞体(cell body)上有一个突触(axon)，它是一个很长的神经纤维，可以传递冲动(impulse)，它的长度从 1mm 到 1 m 不等，蛋白质只能在细胞体中合成，神经末梢所需要的蛋白质必须沿着突触运输过去，这个过程称为突触运输(axonal transport)。蛋白质沿突触运输的速度是比较快的，大约相当于每天 30 cm，或者说每秒 3 μm。蛋白质分子(主要是乙酰胆碱酯酶、膜蛋白和神经纤维(neurofilament 亚基)可以从神经元的细胞体运送到突触的末端，并且可以反向运输。突触运输是依靠突触中的微管进行的，蛋白质由马达蛋白(如动蛋白)帮助沿着微管的表面运行。突触运输受 Ca^{2+}离子的调节。突触运输是目前神经生物学(neurobiology)的重要研究课题。

(五)信号转导

信号转导(signal transduction)是生物的一种重要功能。生物生存的环境中有许多种信号，如光、气味(odorant)、口味(taste)、神经递质(neurotransmitter)、激素(hormane)等，细胞对这些信号非常敏感，能做出迅速的反应。例如，细菌和单细胞藻类遇到营养物质，能朝向营养物而运动，称为趋化性(chemotaxis)。动物的视觉细胞能感受光的信号而做出各种反应，这些过程都属于信号转导。

在细胞表面的生物膜上存在着各种受体(receptor)，它们能与各种化学物质结合形成复合物(complex)。以后受体又与 G-蛋白(G-protein)相互作用，使 G-蛋白的 α、β、γ，亚基解离，G-蛋白的 α 亚基又与细胞内的效应物(effector)相互作用，效应物是各种靶(target)，如酶和离子通道(ion channel)，效应物蛋白能动员第二信使(如 Ca^{2+})，引起细胞内特殊的作用。信号转导过程可以简单地表示如图 3-19。

在信号转导的研究中以哺乳动物眼球的视网膜的杆状细胞(rod cell)的研究最为深入。杆状细胞是细而长的细胞，直径 1 μm，长 40 μm。杆状细胞上存在光受体，即视紫红蛋白(rhodopsin)，它是一种跨膜蛋白，分子量为40 kD，由7股α-螺旋组成，其发色团(chromophore)

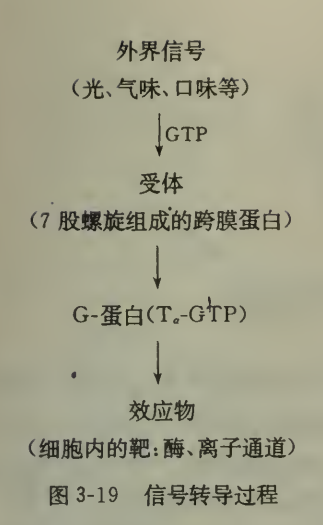


图 3-19　信号转导过程

为 11-cis-视黄醛(retinal)。杆状细胞的转导蛋白(transducin)就是 G-蛋白，它是 GTP 结合蛋白。当光激发视紫红蛋白后，即使 G-蛋白发生构象变化，导致 G-蛋白的解体，α-亚基从 β，γ-亚基上解离下来，随后活化的 α-亚基(T_α-GTP)结合到磷酸二酯酶(phosphodiesterase，PDE)上；可以下式表示：

$$R \xrightarrow{光} R^* \longrightarrow T_\alpha\text{-GTP} \longrightarrow PDE^* \left[\begin{array}{l} \text{cGMP} \longrightarrow \text{张开的通道} \\ 5'\text{-GMP} \rightarrow \text{关闭的通道} \end{array} \right.$$

在视觉刺激的信号转导过程中，信息流是从光激发的视紫红蛋白(R^*)传到转导蛋白(即 G-蛋白，T_α-GTP)，然后再传到磷酸二酯酶(PDE^*)，PDE^* 水解环式 GMP(cyclic GMP)为 5′-GMP，光诱导引起 cGMP 浓度下降，就使阳离子通道关闭。阳离子通道在张开时可使 10 万以上 Na^+ 离子通过，当吸收 1 个光子(photon)后，Na^+ 离子流即被阻断。这个极其显著的放大机制是什么呢？在黑暗中质膜上的阳离子通道靠 cGMP 使之张开。光激发的视紫红蛋白扳动一个开关(switch)，这个开关就是磷酸二酯酶(PDE)，结果就使 cGMP 水解 5′-GMP，阳离子通道乃关闭。实质上，质膜是一个环式 GMP 的电极，可以开闭。

在嗅觉的信号转导中有一个非常类似的离子通道起作用，但是它靠环式 AMP(cAMP)起开关机制。

颜色的识别是通过视网膜圆锥(cone)完成的，圆锥受体有吸收蓝光、绿光及红光三种颜色的受体。圆锥细胞的光受体也是 7 股跨膜蛋白，含有 11-cis-视黄醛作为发色团。位置靠近视黄醛分子的受体，蛋白质氨基酸中的非极性残基被极性残基取代时，其最大吸收峰即从蓝光向红光移动(约 10 nm)。

参考文献

1. 汪堃仁，薛绍白，柳惠图. 细胞生物学. 北京：北京师范大学出版社，1990.
2. Alberts，B.，Bray，D.，Lewis，J.，Raff，M.，Roberts，K. and Watson，J. D. Molecular Biology of the Cell. Garland Publishing，Inc. New York：1994.
3. Lewis，R. Life. WCB，New York：1992
4. 徐是雄(主编). 植物细胞骨架. 北京：科学出版社，1996.
5. Voet，D. and Voet，J. G. Biochemistry Benjamin/Cummings，New York：1990.
6. Darnell，J.，Lodish，H. and Baltimore，D. Molecular Cell Biology. W. H. Freeman，New York：1994.
7. Avers，C. J. Molecular Cell Biology. Addison-Wesley Reading，Massachusetts：1986.
8. 孙大业，郭艳林. 细胞信号系统. 北京：科学出版社，1993.

(阎隆飞)

第四章 DNA 的结构

第一节 DNA 结构的多样性和动态性

“种豆得豆，种瓜得瓜”是古时人们对遗传现象的概括。经过长时期的探索，人们终于获知生物的遗传特性是由一种称为基因的遗传因子决定的，而基因则线性地分布于染色体上，以后又进而证明染色体是由核酸（主要为DNA）和蛋白质组成的。但遗传物质是DNA还是蛋白质的争论仍在继续。1944年，Avery等的肺炎球菌的转化实验指出遗传物质应是DNA；1952年，Hershey和Chase的T2噬菌体对大肠杆菌的感染试验进一步证实了这一点，1953年，Watson和Crick提出了DNA双螺旋结构模型，它使许多遗传现象从分子水平上得到了充分和合理的解释，成为分子生物学发展中的一个里程碑。可是事物总是发展的，人们在以后的研究中，除了证实双螺旋模型的正确性以外，也发现DNA并不是过去所认为的那样是上下一统的棒状分子。DNA双螺旋结构并不是完全规整的，它的构象参数是随核苷酸序列的不同而在一定范围内变化的。以后随之又发现了Z-DNA及其它的稀有结构等。各种实验证据说明DNA结构具有多样性，并且这些多样性的结构还会相互转变，它们处在动态中，这些结果把DNA结构研究推向新的高度。在此同时，基因表达调控的研究显示，DNA序列不仅与编码各种蛋白质有关，也与其调控有关。首先发现的是原核生物一些基因的调控是通过相应蛋白质因子与特定的DNA序列结合实现的，以后又发现真核生物的基因表达虽然更为复杂，但也是通过类似方式实现的，复杂的生命现象需要DNA结构具备相应的多样性，很难以一个统一结构概念来说明DNA的各种功能。

正像蛋白质结构那样，人们也常把DNA结构分为不同的层次——一级结构、二级结构和三级结构。一级结构是指DNA的共价结构和核苷酸顺序，二级结构是指一定或全部核苷酸序列所形成的规律性结构，三级结构是指染色体DNA所具有的复杂折叠状态。当然，蛋白质和DNA的结构层次间并不是完全可以类比的。

第二节 多核苷酸链

多核苷酸链（polynucleotide chain）是核酸（DNA和RNA）结构的基础。它是由核苷酸通过3′,5′-磷酸二酯键连接成的线型大分子。核苷酸是多核苷酸链的结构单位。

核苷酸由三部分组成：戊糖、磷酸和碱基。在核苷酸分子中，戊糖的β-半缩醛羟基与碱基中的氢（嘧啶环1位，嘌呤环9位上的氢）失水缩合成N-苷键而形成核苷，核苷的5′位与磷酸通过酯键形成5′-核苷酸（图4-1），不同的核苷酸用碱基第一个字来命名（表4-1）。核苷酸分子的碱基类似于蛋白质氨基酸的侧链，它们影响着所形成的核酸的结构和功能。

磷酸　　戊糖　　碱　基

核苷　　5′-核苷酸　　胸腺嘧啶　　胞嘧啶　　尿嘧啶(RNA)

图 4-1　核苷和核苷酸

嘧啶和嘌呤都是含氮杂环化合物，具弱碱性，分子所含的闭合共轭体系使其呈现为扁平状。碱基难溶于中性水溶液中，它们的这一疏水特性对 DNA 的结构具有重要意义，碱基的另一重要特性是其环氮及所含的官能团能成为氢键的供体或受体。在 DNA 分子中，一定的碱基通过这些氢键相缔合，这对 DNA 作为遗传物质具有首要意义。

表 4-1　核苷酸的组成和名称

核酸	核苷酸(缩写)	核　苷	碱　基
RNA	腺苷酸(A,AMP)	腺　苷	腺嘌呤
	鸟苷酸(G,GMP)	鸟　苷	鸟嘌呤
	胞苷酸(C,CMP)	胞　苷	胞嘧啶
	尿苷酸(U,UMP)	尿　苷	尿嘧啶
DNA	脱氧腺苷酸(A,dA,dAMP)	脱氧腺苷	腺嘌呤
	脱氧鸟苷酸(G,dG,dGMP)	脱氧鸟苷	鸟嘌呤
	脱氧胞苷酸(C,dC,dCMP)	脱氧胞苷	胞嘧啶
	脱氧胸苷酸(T,dT,dTMP)	脱氧胸苷	胸腺嘧啶

注：核苷酸现已成为包括核糖核苷酸和脱氧核糖核苷酸在内的类名。在核酸分子中，它们都以 A、G、T、C 表示；在游离状态中，脱氧核糖核苷酸则分别以 dA、dG、dT 和 dC 表示。

多核苷酸链是由核苷酸通过 3′,5′-磷酸二酯键连接成的很长的线型分子(图 4-2)。在多核苷酸链中，交替的磷酸和 2-脱氧核糖构成了分子的骨架，碱基为侧链。不同多核苷酸链的骨架是相同的，不同的是碱基的种类和排列顺序。多核苷酸的碱基可简化表示如图 4-2 的右方，也可缩写为 pGpCpTpA 和 pGCTA。按规定，多核苷酸链以连接于 5′-羟基的磷酸开始，以脱氧核糖的 3′-羟基终止。

图 4-2　多核苷酸链片段

在多核苷酸中，磷酸呈四面体构型，脱氧核糖成折叠的 5 元环，而碱基则是扁平的。此外，主链中脱氧核糖的羟基能与水形成氢键，磷酸基在生理条件下离解为负离子，故主链是亲水的，但其侧链（碱基）是疏水的，这些特点规定了多核苷酸链可形成的稳定构象。

第三节　DNA 的双螺旋结构

50 年代初，关于 DNA 的结构可说已积累了不少资料，Chargaff 对 DNA 碱基组成的分析得到了高度规律性的结果：虽然不同来源 DNA 的碱基组成变化很大，但腺嘌呤核苷酸的数目却总与胸腺嘧啶核苷酸相等，鸟嘌呤核苷酸与胞嘧啶核苷酸相等，即 A/T=G/C=1。DNA 纤维的 X 光衍射图指出其中的原子排布沿长轴存在 0.34 和 3.4 nm 两种周期性，这提示分子可能具有螺旋构象。综合这些，Watson 和 Crick 终于在 1953 年提出 DNA 的双螺旋结构模型。他们指出，DNA 是由两条反平行的多核苷酸链组成的，分子的主链位于螺旋的外缘，碱基则堆积于螺旋的内部，两条多核苷酸链通过碱基间的氢键相结合（图 4-3）。

Watson 和 Crick 所提出的 DNA 双螺旋结构的特征是：

构成 DNA 的两条多核苷酸链的磷酸-戊糖骨架盘旋成直径为 2 nm 的双螺旋，由于脱氧核糖中连接碱基的 C1′并不正好处在螺旋的相对位置上（见图 4-3），故两股链在螺旋轴上的间距并不相等，从而在分子表面形成宽窄不等的大沟（major groove）和小沟（minor groove）。双股多核苷酸链上的碱基则位于螺旋的内部，大沟垂直于螺旋轴，相邻碱基的间隔为 0.34 nm；每一匝螺旋含 10 bp（以后从溶液中测得的数字为 10.5），其轴向距离为 3.4 nm。每一匝螺旋转动 360°，故相邻碱基对间转动 36°。

双螺旋中，碱基的配对关系是严格的，即腺嘌呤和胸腺嘧啶配对，鸟嘌呤和胞嘧啶配对，碱基的这种严格配对关系基于两方面的因素：一是形成氢键的基团在空间的取向；二是碱基对的大小能容纳于双螺旋中。由于双股链是反平行的，其立体化学取向正好相反，这就使双螺旋的外形，无论从何端观察都相同，呈现出二重对称性。碱基的上述配对关系称为 Watson-Crick 配对（图 4-4）。

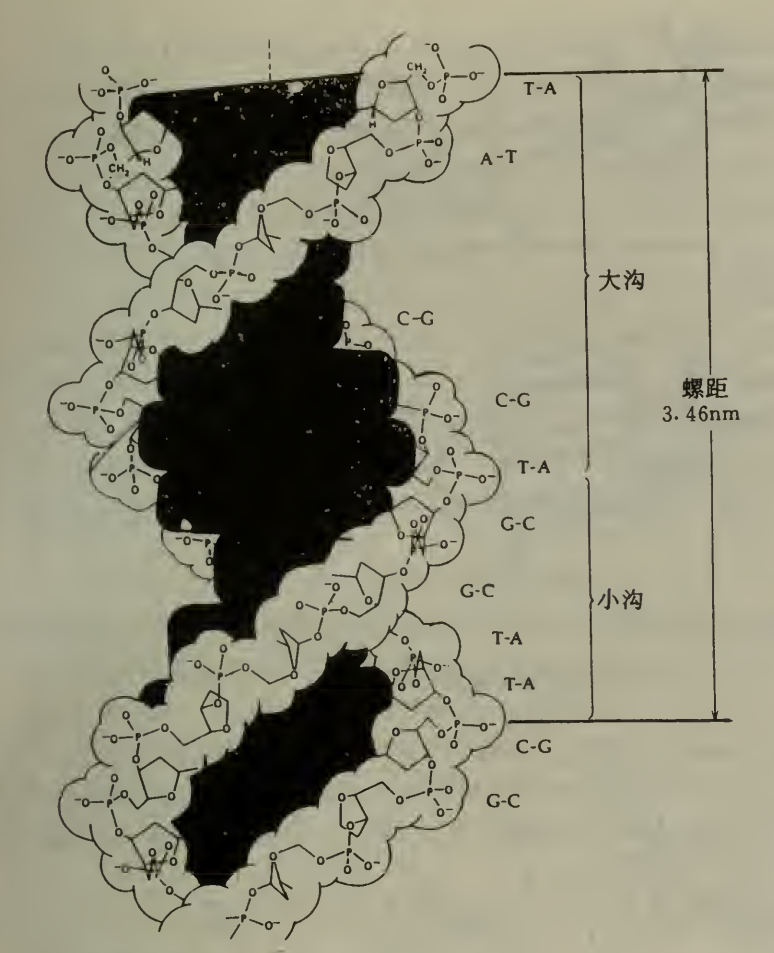
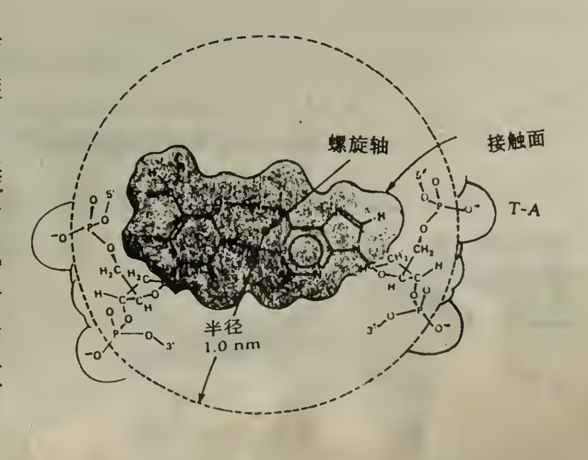


图4-3 DNA 的双螺旋结构

左图：示大沟小沟；

下图：示碱基对。



作为遗传物质，DNA 的结构是稳定的。多核苷酸链的疏水交互作用是形成双螺旋结构的一个重要因素，能力学上它有利于螺旋结构的稳定。在双螺旋结构中，碱基对上下间隔0.34 nm，这相当于苯的厚度，并表明碱基的堆积是紧密的，碱基间存在着包括 van der Waals 力在内的堆积力。碱基在横向则由氢键相结合。此外，多核苷酸链骨架上的氧原子还能与水形成氢键。这些因素都使 DNA 结构趋向稳定化。双螺旋骨架中磷酸基的氧原子带负电荷，其相斥作用会降低螺旋结构的稳定性。在细胞中，DNA 与碱性的组蛋白、亚精胺以及 Mg^{2+} 等结合，这抵销了负电荷间排斥作用。

DNA 是遗传信息的载体，它的核苷酸序列不仅编码各种蛋白质，并且也与基因表达的调控有关，即它们能为一些蛋白质因子所识别和结合。现在认为这一过程主要是通过大沟实现的，因为大沟比小沟携带更多的形成氢键的信息（图4-5）。图中的圆圈代表螺旋的截面，大于

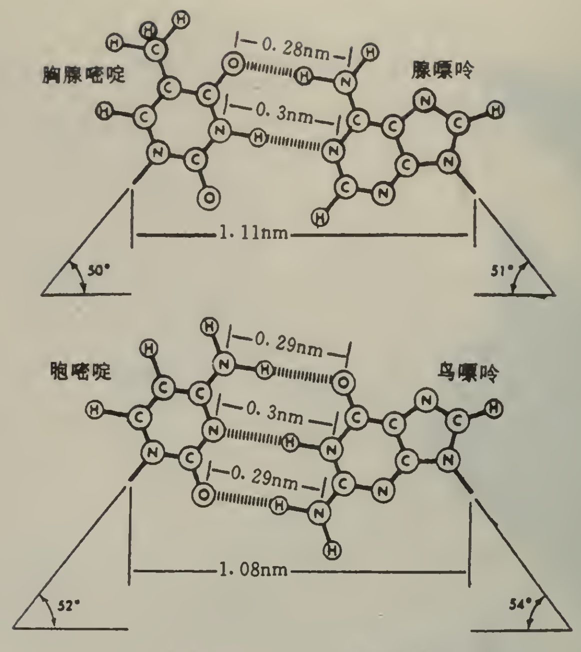


图 4-4　DNA 中的碱基配对

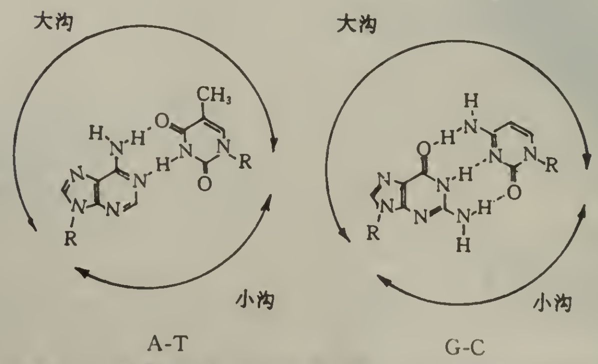


图 4-5　碱基的可能形成氢键的基团在大沟小沟中的分布

180°的一侧表示大沟，另一侧为小沟。在大沟中裸露的是碱基对顶部的基团，不同的碱基对这些基团排列的次序不同，从而也就制约了与之结合的蛋白质因子的结构。这些基团在 4 种碱基对中的顺序为：

大沟：		小沟：	
A-T	环氮，氨基，酮基	A-T	环氮，　酮基
T-A	酮基，氨基，环氮	T-A	酮基，　环氮
G-C	环氮，酮基，氨基	G-C	环氮，氨基，酮基
C-G	氨基，酮基，环氮	C-G	酮基，氨基，环氮

从上不难看出小沟中显示出的信息要少些。蛋白质因子沿 DNA 大沟形成氢键的可能性较沿小沟更多地依赖于碱基序列，即专一性更强。但小沟对 DNA 的功能也不是不重要的(见第 11 节)。

DNA 双螺旋结构的建立对分子生物学的发展无疑意义重大，但也不是从未受到挑战。

DNA 具有双螺旋结构的最初证据来自 DNA 纤维的 X 射线衍射实验，但这一方法在当时的条件下还难于确立 DNA 结构的细节；以后，分子模型的建造回答了多核苷酸链根据其化学特征可以折叠成怎样的三维结构，并当腺嘌呤和胸腺嘧啶及鸟嘌呤与胞嘧啶的配对关系一旦建立时，一种与立体化学相符的规整双螺旋分子就出现了。这一模型不仅可以解释 DNA 的碱基组成，并且也通过多核苷酸链上碱基间的互补关系为 DNA 的复制提出了一种清晰的机制。有理由可以认为双螺旋结构模型是正确的。

不过，多年积累起来的有关 DNA 结构的实验资料只可能认为与通过分子模型建立起来的双螺旋结构相一致。因为 DNA 的序列的高度异质性使所得到的 DNA 纤维的取向难以达到能提供各个原子定位所需的精确度，这使人们不能像蛋白质（例如血红蛋白）那样藉 X 射线衍射法来建立确定的分子结构。因此，长时间来虽没有分子生物学家对碱基配对的双螺旋模型提出重大的怀疑，但亦无法确证 DNA 是右手性的和它的碱基配对确是通过氢键结合的。后来，有机化学家终于合成出了具确定核苷酸序列的寡核苷酸晶体，例如：

GGTATACC

CCATATGG

X 射线衍射分析证明，它确是长期来所认为的通过碱基间形成氢键而配对的双螺旋结构。

DNA 的结构最近由于应用扫描隧道显微镜（scanning tunnel microscope，STM）而得到了进一步的证实。扫描隧道显微镜是近几年发展起来的一种用于研究物质表面的新设备。它利用量子隧道效应原理，将原子线度的极细针尖在接近样品处（<1 nm）扫描，通过监测样品与针尖间隧道电流随距离的变化从而获知样品表面的面貌。现在，人们通过 STM 已直接观察到天然 DNA 的双螺旋结构，STM 正在成为研究核酸、蛋白质结构的新的有力工具。

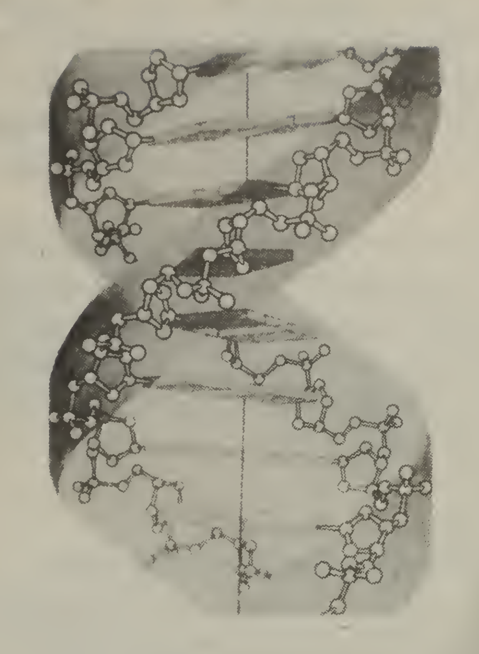
第四节　A-DNA 和 B-DNA 结构的多态性

在溶液中，DNA 一般为 B 型，但当水合的 B-DNA 脱水，或由于加人乙醇或盐而使水的活度降低时就转变为 A 型。B-DNA 中，2-脱氧核糖的构象出现了变化从而使同股多核苷酸链中相邻磷酸基间的距离缩短 0.1 nm。这一变化使每匝螺旋的碱基对数由 B-DNA 的 10（溶液中为 10.5）转变为 A-DNA 的 11。A-DNA 的另一重要差别在于碱基对在螺旋中的位置，在 B-DNA 中，碱基对集中于螺旋轴；在 A-DNA 中，碱基对向大沟方向移动了约 0.5 nm，这一变化的结果改变了 B-DNA 的外形和沟的尺寸（图 4-6）。

在实验室中，从互补的寡核苷酸获得的 DNA 晶体由于制备过程中化学试剂的作用而呈现为 A-DNA。不过，现在还不清楚，细胞中是否存在纯的 A-DNA 片段。在转录过程中，在 DNA 单股上合成 RNA 时，DNA 与所合成的 RNA 形成杂交双螺旋，这种片段可能是 A 型，这是由于 RNA 中核糖的 C2′ 羟基强烈趋向于采取 C3′-内式构象。现已考查过的 RNA 双螺旋也采取 A 型构象。

除了 A-DNA，B-DNA 外，人们也还观察到了其它的、结构参数有一定差异的双螺旋 DNA。这一现象称为 DNA 结构的多态性（polymorphism）。即将在下一节中讨论的 Z-DNA 即是其中之一。

DNA 结构产生多态性的原因在于多核苷酸链的骨架含许多可转动的单键，从而使糖环可采取不同的折叠形式和苷键采取不同构象。



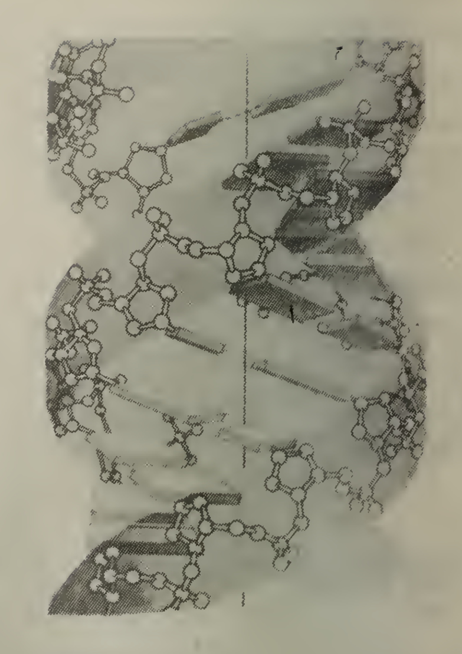


图 4-6　B-DNA 和 A-DNA
右侧为 B-DNA，左侧为 A-DNA。

多核苷酸链骨架含有许多可转动的键(图 4-7)，但实际上主要的转动发生在磷酸二酯键的两个 O—P 键上。多核苷酸链的 N-苷键也可转动，从而使糖环和碱基处在不同的空间关系中，两个极端的情况是反式和顺式的构象(图 4-8)。在反式构象中，嘧啶碱基的 CO 基(或嘌呤碱基的嘧啶部分)远离糖环；在顺式构象中，上述基团位于糖环上方。在 DNA 中，反式较顺式构象远为多见。

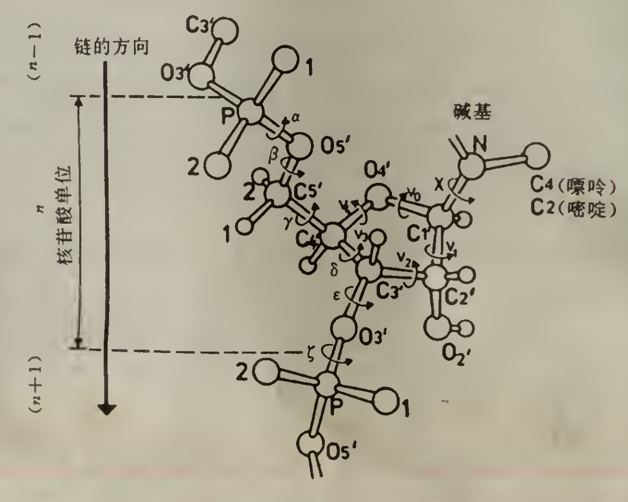


图 4-7　多核苷酸链中的可转动键

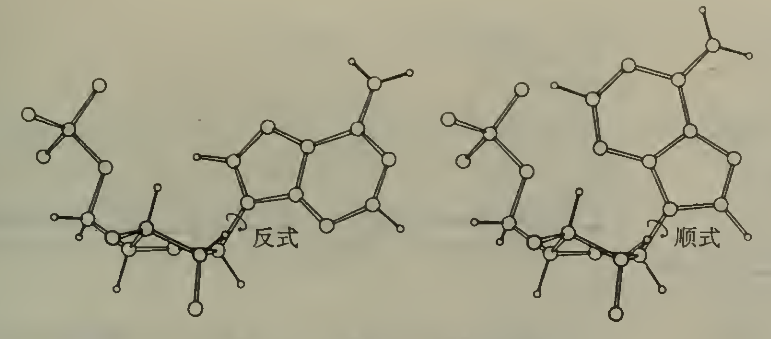
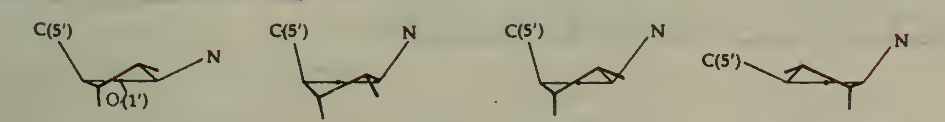


图 4-8　苷键的反式和顺式构象

DNA 中的脱氧核糖为呋喃型，它的环状结构由 4 个碳和 1 个氧原子构成，这 5 个原子不是共平面的，而形成不同的折叠形式。图 4-9 列出的是常见的几种，其中 c 为中间类型，易于转变成其它三种。在 a，b 和 d 中，氧和 3 个碳原子处于同一平面中，C2′ 或 C3′ 位于平面的上方或下方。C2′ 或 C3′ 位于 C5′ 同侧的构象称为内式，位于相反一侧的称为外式。



a. C(2′)-内式　b. C(3′)-外式　c. C(2′)-内式-C(3′)-外式　d. C(3′)-内式

图 4-9　脱氧核糖的一些构象

由于上述因素的差异，具同样序列的 DNA 即呈现出构象上的多态性(表 4-2)。

A-DNA，B-DNA 和 Z-DNA 的一些构象参数有所不同(表 4-2)，其外形也呈现出差异(图 4-12)。

表 4-2　A-DNA，B-DNA 和 Z-DNA 的一些结构特性

	A-DNA	B-DNA	Z-DNA
螺旋方向	右手性	右手性	左手性
每圈螺旋的碱基数	11	10	12
每一碱基对的上升距离	0.255 nm	0.34 nm	0.37 nm
螺　距	2.8 nm	3.4 nm	4.5 nm
碱基对的倾斜度	20°	6°	7°
每一碱基对在螺旋中旋转的角度	33°	36°	−60°(每一个二聚体)
糖苷键的构象			
脱氧胞嘧啶核苷	反	反	反
脱氧鸟嘌呤核苷	反	反	顺
糖的折叠			
脱氧胞嘧啶核苷	C′-3-内式	C′-2-内式	C′-2-内式
脱氧鸟嘌呤核苷	C′-3-内式	C′-2-内式	C-3-内式
总尺寸	短而宽	较长而细	长而细
沟的尺寸			
大沟	细而深	宽和中等深	平伏于螺旋表面
小沟	宽而浅	窄和中等深	很窄和很深

第五节 Z-DNA

在 DNA 双螺旋构象中，最使人感到震惊的是 Z-DNA。Z-DNA 是 Rich 及王惠君等对寡聚核苷酸 d(CpGpCpGpCpG)进行 X 射线衍射分析后发现的。Z-DNA 不同于 B-DNA，它是左手双螺旋。它的基本重复单位为嘌呤-嘧啶二核苷酸，其中嘧啶核苷酸中的苷键为反式取向，而嘌呤核苷酸的苷键为顺式取向。在右手双螺旋中，苷键的构象都是反式。这表明嘌呤核苷酸苷键构型的不同产生了左手双螺旋。

与 B-DNA 相比，Z-DNA 更为细长，Z-DNA 中，每一碱基对在螺旋轴上的上升距离为 0.37 nm，螺距为 4.5 nm，每一螺旋圈含 12 bp。与 B-DNA 比较，Z-DNA 中的碱基对平面相对于螺旋轴转动了 180°(图 4-10)。碱基对平面的翻转对交替 d(GC)序列中两种残基的构象产生不同的影响。在这一过程中，胞嘧啶核苷残基整个转动了 180°，因而仍保持反式构象，而鸟嘌呤核苷残基中的碱基则绕苷键转动了 180°，其结果就使多脱氧核苷酸主链的走向呈现为锯齿形(图 4-11)。产生这一差异的原因已由分子模型得到说明：核苷的反式构象所受到的空间挤压虽然较轻，但 dG 和 dC 比较，后者难于采取顺式构象。

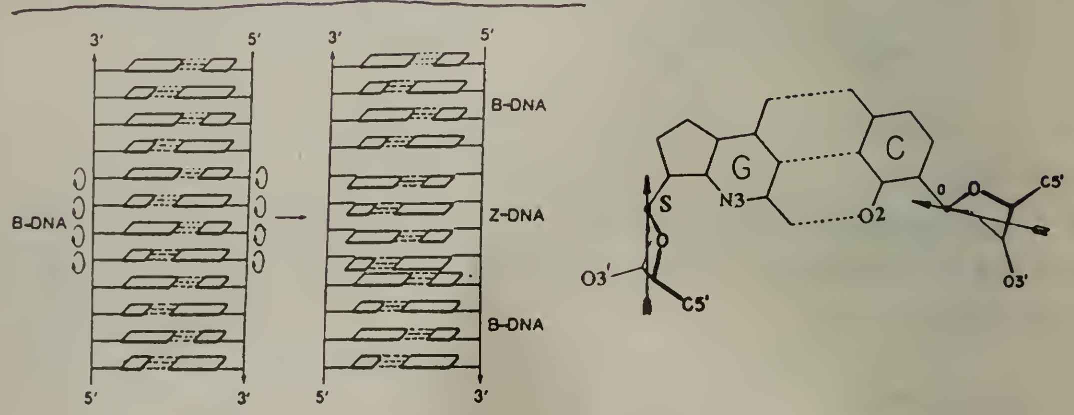


图 4-10　B-DNA 向 Z-DNA 转变中碱基对平面的翻转　　　图 4-11　Z-DNA 中 d(GC)残基的构象

Z-DNA 中，G-C 碱基对并非对称地位于螺旋轴附近，它们向边缘伸展，从而使大沟外凸，小沟则变得窄而深(图 4-12)。

Z-DNA 的糖-磷酸主链的伸展程度不如 A-或 B-DNA。在溶液中，只有正离子(如 Na^+)的浓度高至足以中和磷酸基上的负电荷时，交替的嘌呤-嘧啶才采取左手性构象。盐浓度较低时，它形成典型的 B-DNA。

由于 GC 对远较 AT 对为稳定，因此，含交替 d(GC)的双螺旋短片段比相应的 d(AT)片段更易形成晶体。最早的左手性 DNA 晶体就是由交替 d(GC)寡核苷酸获得的；随后，由更为复杂的序列如：

CGCATGCG
GCGTACGC

也得到了完美的 Z-DNA 晶体。现在认为，在适当的离子条件下，任何含不少于 6 bp 的交替嘌呤-嘧啶序列都能形成 Z 型片段。以后，1985 年，X 射线衍射法进一步证明如下的序列：

^{m5}C　G AT^{m5}C　G

　G^{m5}C TA　G^{m5}C　　(m 指甲基化)

也能形成 Z-DNA 晶体。这说明,如胞嘧啶残基的 5′位一经甲基化(由专一性的甲基化酶催化),交替嘌呤-嘧啶序列就不再构成产生 Z-DNA 的必要条件。我们现在还不可能仅靠对核苷酸序列的考察就正确地指出可能形成 Z-DNA 的区段。

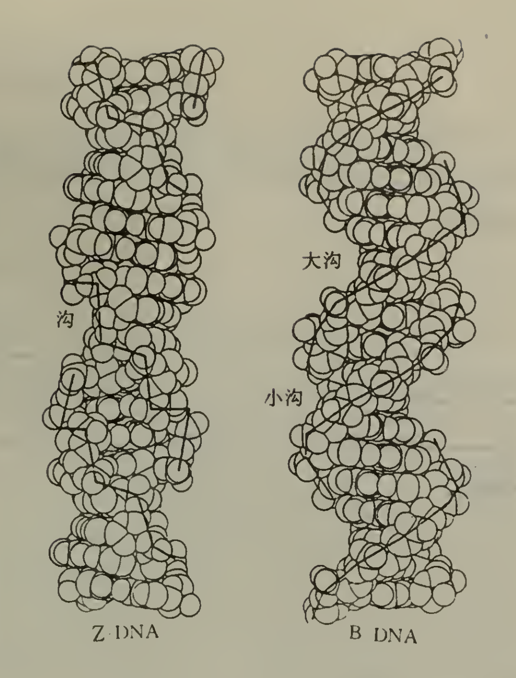


图 4-12　B-DNA 和 Z-DNA

现在已有证据说明 Z-DNA 存在于天然 DNA 中。应用针对 Z-DNA 的抗体已可鉴别出天然 DNA 中的 Z-DNA 区段。但细胞中所存在的 Z-DNA 的量至今仍不清楚。对此的一个最可能的估计是,细胞中的 DNA 只有微小的区段成为 Z 型。Z-DNA 在天然 DNA 中的存在表明它应有自己的独特功能。已有一些证据表明 Z-DNA 的存在与基因表达的调控有关。

现在已发现大多数染色体 DNA 中的少部分胞嘧啶残基的 C5 是甲基化的,这是在胞嘧啶残基掺入 DNA 后由专一性的 DNA 甲基化酶催化的。同样,DNA 中的腺嘌呤也可通过另一种专一性甲基化酶的作用转变成 6-甲基腺嘌呤。在原核细胞中,甲基化腺嘌呤较甲基化胞嘧啶更为多见,而在真核细胞中,甲基化的几乎都是胞嘧啶残基。

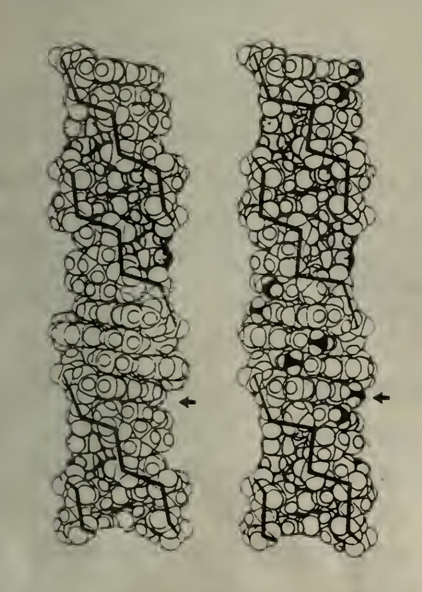


图 4-13　(dG-dG)$_2$和(^{m}dC-dG)$_2$ 的结构
左方为未甲基化的 Z-DNA,右方为甲基化的 Z-DNA。图中的黑线表示主链,黑色图形表示甲基中的碳原子,箭头标出的是 Z-DNA 中的凹陷部。

在真核细胞中,DNA 甲基化酶优先修饰其 3′侧为鸟嘌呤的胞嘧啶(5′CG3′)。但并不是所有的 5′CG3′中的 C 都是甲基化的,在多细胞生物的不同类型细胞中,特定基因的甲基化程度也相差很大。有迹象表明,胞嘧啶残基的甲基化可能与基因表达的调控有关。

关键性胞嘧啶残基的甲基化可能通过使 B-DNA 转变为 Z-DNA 而影响基因功能的发挥。在大多数细胞的阳离子条件下,交替的 CG 区段很可能以 B 型存在,但在胞嘧啶被甲基化后,平衡就转向有利于 Z 型。产生这一转变的理由是,在 B-DNA 中甲基突出于大沟的含水环境中,这是一种不稳定状态;在 Z-DNA 中,相同的甲基转而形成了较稳定的疏水区(图 4-13)。

第六节 DNA 的变性和复性

双股链的互补是 DNA 在结构和功能上的一个基本特性，也是 DNA 研究中一些实验技术的基础。

一、DNA 的变性

DNA 的生物学功能依赖于其结构。当双螺旋 DNA 加热至生理温度以上（接近 100℃）时，它就失去生理活性。这时 DNA 双股链间的氢键断裂，最后双股链完全分开并成为无规线团。这一过程叫做 DNA 的变性（denaturation）或融解。DNA 的变性发生在一个狭窄的温度范围内，并伴随着许多物理性质的变化。这些物理性质的变化中，特别有用的是增色效应（hyperchromatic effect）。在 DNA 的变性过程中，它在 260 nm 的吸收值先是缓慢上升，到达某一温度后即骤然上升（图 4-14），故称为增色效应。天然 DNA 和变性 DNA 的吸收值相差可达 34%。吸收值的骤然上升表明 DNA 的变性是一协同性过程。吸收值增加的中点称为融解温度（melting temperature，T_m）。在接近生理条件下，DNA 的 T_m 一般在 85～95℃间。DNA 的 T_m 为多种因素所影响。T_m 决定于 DNA 的 G+C 含量；G+C 含量高，T_m 也高（图 4-15）。溶液的离子强度也会影响 T_m，因为离子强度会影响 DNA 的稳定性，每升 0.18 mol Na^+ 是常用的标准条件。T_m 还会因甲酰胺、尿素等的存在而降低。这是由于甲酰胺等会增加碱基在水中的溶解度，从而减弱碱基的疏水交互作用。当溶液的 pH 接近 12 时，DNA 碱基的酮基会转变成烯醇基；当 pH 接近 2～3 时，碱基上的氨基即质子化。pH 的这些变化都会影响碱基间的氢键的稳定性从而改变 T_m。故当这些条件发生变化时，DNA 的 T_m 也就改变。例如，在甲酰胺的存在下，T_m 可降低至 40℃。又如在中等温度或中等量的碱、甲酰胺等去稳定剂存在下，DNA 会部分变性。在这种部分变性的分子中，富于 AT 的区域融解，而富于 GC 的区域仍保持其双螺旋特性。

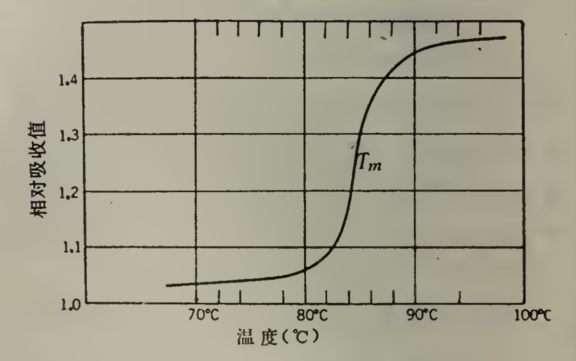


图 4-14 DNA 的融解曲线

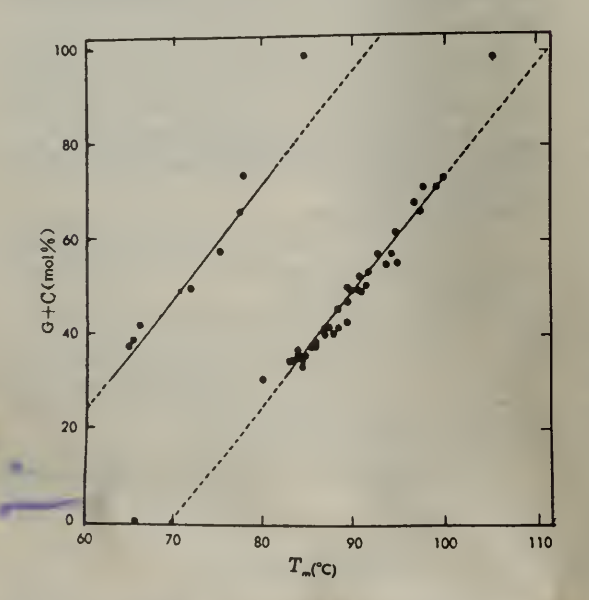


图 4-15 两种不同溶液（pH7）中，DNA T_m 对 G+C 含量的依赖性

二、DNA 的复性

热变性的 DNA 如缓缓冷却，已分开的互补链又可能重新缔合成双螺旋，这叫做复性(renaturation)。复性的速度与 DNA 的浓度有关，因为两互补序列间的配对决定于它们的碰撞频率。对高分子量 DNA 言，这一过程更为复杂，因为分子会由于部分互补而形成不完全的双螺旋，这时在 260 nm 的吸收值虽会有所恢复，但不会恢复至天然 DNA 溶液的数值。如使 DNA 在剪切力作用下分割成较短的片段，它们就可完全复性。DNA 复性的速率还可用为量度其序列复杂性(complexity)的方法。两种含相同缓冲剂的变性 DNA 溶液，如它们的浓度相等(即单位容积中含相同的核苷酸数)而来源不同(例如一种来自小病毒；另一种来自脊椎动物)，则小病毒的复性就要快得多。

DNA 的变性和复性是可逆的。如 DNA 仅部分变性，它的复性就表现在快速的一级反应，如 DNA 已完全变性，它就表现为缓慢的二级反应。不完全变性 DNA 的复性只是已分开部分的双股链间碱基的重新配对，它犹如拉链的扣合。完全变性 DNA 的复性就不同，两条完全分开的互补链只有在正确位点上形成氢键才能导致全面复性。既然多核苷酸链是由许多核苷酸组成的，因而两者在正确位点上的缔合就是一个不断尝试的过程，两股链的浓度越高，成功的频率也高，故它的复性就表现为二级反应。现如以 C_o 代表完全变性 DNA 生成的单股链的起始浓度，则在时间 t 时的浓度变化就是：

$$\frac{\mathrm{d}\,C}{\mathrm{d}\,t}=k\,C_o^2$$

积分后得

$$\frac{C}{C_o}=\frac{1}{1+kC_ot}$$

上式中的 k 为复性的速度常数。以 C/C_o 对 C_ot 作图即得 C_ot 曲线(图 4-16)。当 $C/C_o=0.5$ 时的 C_ot 值称为 $C_ot_{½}$。$C_ot_{½}$ 与 k 的关系为：

$$0.5=\frac{1}{1+k\,C_ot};\quad k=\frac{1}{C_ot_{½}}$$

k 对 DNA 具有特征性，它与 DNA 的碱基对数目(N)成反相关，因此，C_ot 曲线提供了一种测定 DNA 分子量的方法。C_ot 曲线也揭示单一来源的 DNA 所具有的不同复杂性部分。图 4-16(a)中的①和②分别为 MS2 噬菌体和 T4 噬菌体 DNA 的 C_ot 曲线；(b)中③为小鼠 DNA 的 C_ot 曲线。

基因组的复杂性是用 DNA 中单一序列的长度加上各个重复序列的长度来表示。如某一基因组含序列 a，b，c，d 而无重复，则其复杂性即为 $a+b+c+d$。这时复杂性即与分子量相同。高等生物的 DNA，除单一序列外还常含重复序列。例如，某一生物的 DNA 由 10^6 拷贝的 a，10^3 拷贝的 b，10 拷贝的 c 和 1 拷贝的 d，e 和 f 组成，其复杂性为 $a+b+c+d+e+f$，这时 DNA 的复杂性就大大小于其分子量。复杂性的单位为核苷酸对的数目。在图 4-16 中，①的复杂性低于②，③的复杂性最高，它由 3 个动力学成分组成。复性动力学(renaturation kinetic)是研究 DNA 复杂性的一种有用方法。

应该指出的是，图中的单一动力学成分是一平均值，它可以是由几种序列组成的，它们之所以会形成一个独立成分是在于它们在总 DNA 中占有明显的比例。

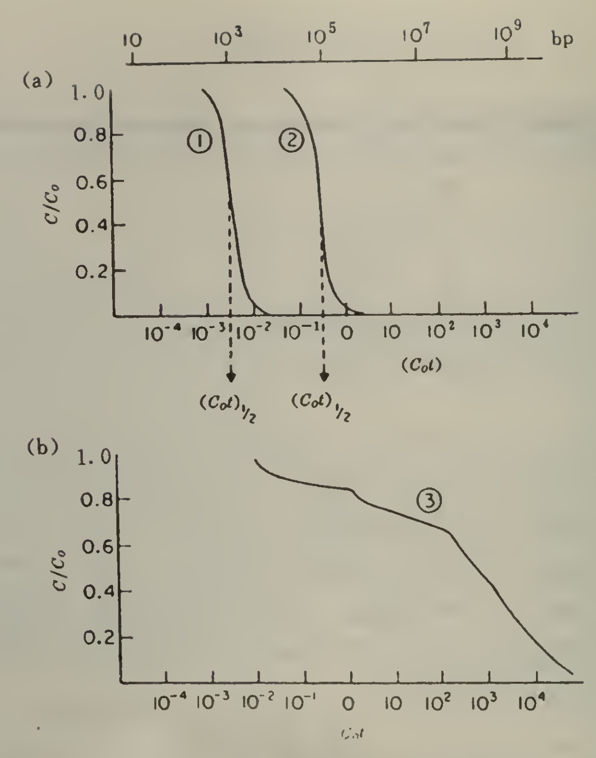


图 4-16　3 种不同来源 DNA 的 C_ot 曲线

图上方的标尺为标准条件下的序列复杂性，横坐标表示的是 C_ot。

三、分子杂交

由 DNA 复性研究发展成的一种实验技术是分子杂交。杂交可以发生在 DNA 和 DNA 或 RNA 间。DNA：DNA 杂交可用以估测 DNA 间的同源序列，不同生物在进化过程中的相关性，以及 DNA 限制片段的作图等。DNA：RNA 杂交可通过 RNA 转录本检测 DNA 中特定基因的存在等。

DNA：RNA 杂交中的一个应予充分考虑的问题是 DNA：RNA 与 DNA：DNA 杂交间的竞争，从而避免杂交片段的漏检。这一困难是可以克服的，因为 DNA：RNA 杂交分子更为稳定，故采用一个稍低于其 T_m 的温度可使其优先形成。此外，这一困难也可通过不同RNA 浓度并使其双倒数图外插到无限浓度时的杂交量来得到进一步解决（图 4-17 a，b）。这一操作常采用放射性 RNA，未杂交的 RNA 用 RNA 酶 A 除去，杂交分子可先沉淀然后计数，或是用硝酸纤维薄膜分离。

有些工作也可不测定所形成的杂交分子，而是测定在一种核酸大量过剩条件下杂交分子的形成速率（图 4-17c，d）。这一方法类似于 C_ot 法，当过剩的为 RNA 时称 R_ot 法，当过剩的为 DNA 时称 D_ot 法。在 R_ot 法中，杂交分子的形成速率为 R_ot（R_o 为 RNA 的总浓度）的函数；在 D_ot 法中，杂交分子的形成速率为 D_ot（D_o 为 DNA 的总浓度）的函数。杂交分子的形成速率在 R_ot 法中，以 DNA 的放射性不再为核酸酶 S1 所改变来表示。在 D_ot 法中，以 RNA 的放射性不再为 RNA 酶 A 所改变来表示，图 4-17 的（c）指出样品中存在两类 RNA 转录本，其中①所含的拷贝数较②为多；（d）指出 DNA 存在三类同源序列，其中③含的拷贝数最多。

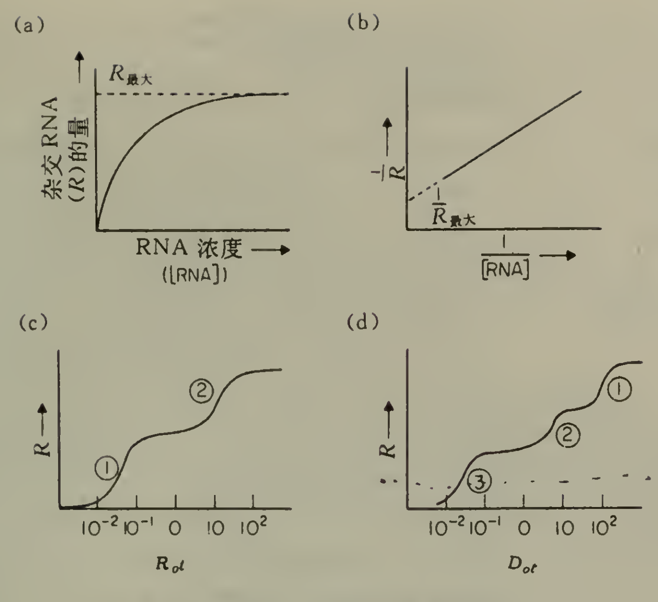


图 4-17　RNA : DNA 分子杂交曲线

R 表示已杂交的 RNA 量。

第七节　DNA 的形状、大小和序列组织

一、DNA 的形状和大小

DNA 一般为长而无分支的双股(ds)线型分子,但有些为环型,也有少数为单股(ss)环型。不同的 DNA,大小相差悬殊。SV40 含 5.1 kb,而南美肺鱼的基因组含102 000 000 kb。虽然一般言,复杂的机体需要更多的 DNA,但不存在严格的对应关系(表 4-3)。

二、DNA 的碱基组成和序列组织

在所有生物种的双股 DNA 中,A 和 T 的量相等,G 和 C 的量相等。故根据 A,T,G,C 中任一种的含量就可确定双股 DNA 的其它 3 种碱基的含量。DNA 的碱基组成一般可用其 GC 含量来表示。

同一物种所有细胞 DNA 的 GC 含量都相同,而种间的差异则很显著,尤其在细菌间,可从 0.3～0.7。高等生物的 GC 含量一般小于 0.5。

DNA 中,除了 A,T,G,C 外还含有修饰碱基,常见的为甲基化碱基,它们是在 DNA 合成后形成的。在大多数动物和植物中,胞嘧啶,尤其是 CpG 中 C 的 5 位常具有甲基,它们与发育过程中基因的钝化有关。在细菌中也已发现碱基的修饰作用,它可保护自身 DNA 免被细胞内的限制酶所降解。在大肠杆菌和相关的细菌中,转座酶的启动子含 GATC 序列,其中 A 能为甲基化酶所甲基化。当双股链中 GATC 序列中的 A 都被甲基化时,启动子相对钝化,复制后形成的半甲基化 DNA 中,启动子是较为活跃的。这是细胞在复制过程中控制转座的方式之一。

表 4-3　DNA 分子的形状和大小

机　体	基因组大小		类　型
	碱基对数(kb)	总长度(μm)	
病毒			
SV40	5.1	1.7	环型
ΦX174	5.4	1.8	单股环型
M13(fd,f1)	6.4	2.1	单股环型
P4	11.6	3.6	线型
T7	39.9	13	线型
P22	43.2	14	线型
λ	48.6	16	线型
T2,T4,T6	166	55	线型
牛痘	190	63	线型
禽痘	280	93	线型
细菌			
类菌质体	760	260	环型
大肠杆菌	4 700	1 360	环型
真核生物			
酵母	13 500	4 600	16 单倍染色体数
果蝇	165 000	56 000	4 单倍染色体数
人	2 900 000	990 000	23 单倍染色体数
南美肺鱼	102 000 000	34 700 000	19 单倍染色体数

原核生物 DNA 的碱基组成是均匀的，即使 DNA 被破碎为较小的，例如，10 kb 片段后，氯化铯密度梯度仍只能使其形成单一的峰。可是真核生物 DNA 的碱基组成与此相反，显示出很大的不均匀性。这反映为 DNA 复性曲线的复杂性和其剪切片段在浮力密度梯度中的不对称分布(图 4-18)。这两方面的实验证据都表明 DNA 中存在着一些独特的序列。

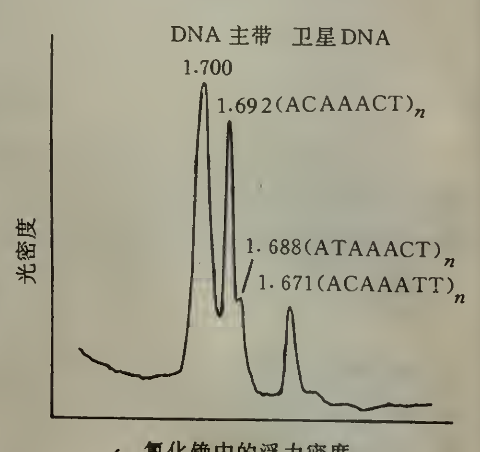


图 4-18　果蝇的卫星 DNA

真核生物 DNA 碱基组成上的异质性主要由于存在着 3 类 DNA 序列：①高度重复序列；②中度重复序列和③单一序列。这 3 类序列各具特点和意义。

(一)高度重复序列

高度重复序列(highly repetitive sequence)的长度为 2～10 bp，在基因组中可串联重复至 10^5～10^7 次，约占基因组的 1%～50%，平均为 15%。一些蟹的重复序列为交替 A,T,G 和 C，仅占 3%并散布其间。大鼠和果蝇的重复序列分别由 6 个和 7 个碱基组成。这些重复序列常称为卫星 DNA(satellite DNA)。

蟹　　A—T —A—T—A—T

大鼠　G—C —A—C—A—C—G—C—A—C—A—C

果蝇　A—C —A—A—A—T　T—A—C—A—A—A—T—T

同一物种的不同卫星序列谱带常是相关的，例如，在果蝇的一个种中发现的重复 10^7 次的 3 种卫星序列为：

5′-ACAAATT-3′

5′-ACAAACT-3′

5′-ATAAACT-3′

有些卫星 DNA 的序列很长，例如，猿猴 DNA 的重复序列长达 172 bp，一种果蝇重复序列长达 359 bp，而牛的卫星 DNA 含 1 400 bp 重复单位。

高度重复序列的功能和确切位置都不完全清楚，它们一般不转录。现在已知，有些卫星 DNA 位于染色体的中心粒，它们可能在减数分裂中承担一定的功能。也有一些重复序列位于染色体的端粒，它们是由一些富含 GC 的序列组成，重复 20～100 次。端粒的功能在于保持染色体的稳定(端粒的结构将在第十二节中进一步讨论)。

高度重复序列也散布于整个基因组，它们可分为长(6～7 kb)和短(79～300 bp)散布重复序列两类，它们是由细胞中的 RNA 通过逆转录形成并插入染色体个别部位的。研究得最透彻的短散布序列是灵长类 DNA 中的 Alu 家族(Alu family)。它是根据所含的限制性内切核酸酶 Alu I 的识别位点而命名的。Alu 家族含一种 281 bp 序列的各种变体，人类基因组含有500 000 拷贝，构成总 DNA 的 5%～6%。Alu 家族原来是由 7 SL RNA(一种信号识别颗粒 SRP 的组成 RNA)逆转录而来。在其它生物体中也已发现相关的序列，如小鼠的 B1 家族，它的序列与 Alu 家族相似而长度仅为其一半。中国仓鼠也含类似的序列，大多数的短散布重复序列不是由 7 SL RNA 而是由转移 RNA 逆转录来的。

这些散布重复序列大致是在进化的较近阶段中形成的。

(二)中度重复序列

中度重复序列(moderately repetitive sequence)包括 rRNA 和组蛋白基因，每一基因组约含 1 000 拷贝。核糖体 DNA 序列不仅在真核基因组中重复许多次，即便原核 DNA 也含多重拷贝。基因的这种重复的目的在于满足蛋白质合成对核糖体的需求。南非爪蟾的 rRNA 基因存在于卫星带中，18 S 和 28 S RNA 基因成串联重复，并为高 GC 含量的间隔区所分开。其它的中度重复序列不是串联而是分布在整个基因组中的。

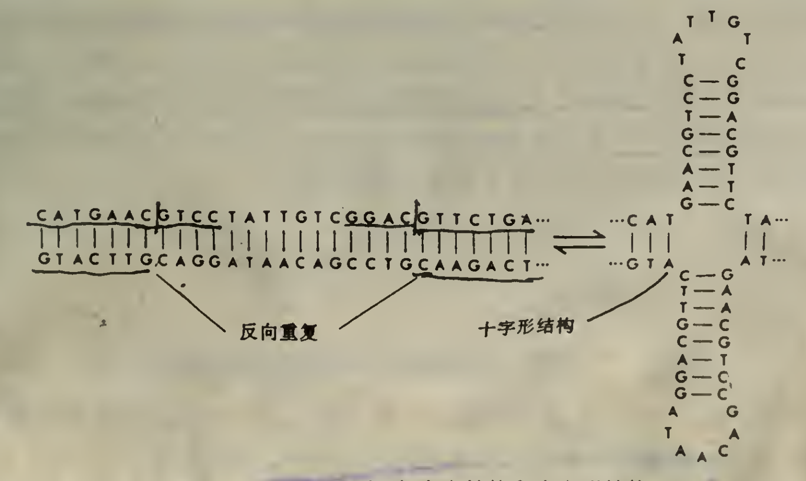


图 4-19　反向重复序列、发夹结构和十字形结构

(三)单一序列

单一序列(unique sequence)包括酶在内的各种蛋白质基因，数目在一个到几个。单一序列

构成人基因组的 65%和约接近细菌、病毒以及质粒基因组的全部。

(四)反向重复序列

反向重复序列(inverted repetitive sequence)是 DNA 中的一种常见的序列形式(图 4-19)。包括一种能与同股或另一股链中相应互补的序列。这种序列单股时形成发夹结构(hairpin structure),双股时形成十字形结构(cruciform)。这些序列能迅速复性。在果蝇、爪蟾和人 DNA 中发现的这种序列长度达 1 200 bp。四膜虫大核(macronucleus)的核糖体 DNA 位于 20 kb 的反向重复序列中。发夹序列约占人基因组的 6%,在单倍染色体中的数目约为 2×10^6。这一状况从未在原核生物中发现。

三、真核和原核生物基因组的进一步比较

真核和原核生物基因组的序列组织(sequence organization)已在上节做了比较,但两者的差别不仅表现在重复序列上,真核生物 DNA 的单一序列也具不少特点:

(1)大多数的蛋白质基因(包括部分 RNA 基因)是割(断)裂基因(split gene),它们含有一些称为内含子(intron)的、转录后被切除的序列,内含子常使基因扩大 20 倍。不同内含子的长度和序列是有差异的,它们可以存在于基因的各个部位。至今,内含子是否具有何种功能仍是一个正在探索中的问题。

(2)真核生物 DNA 的基因间存在着转录或不转录的间隔区,即使是转录的间隔区亦不翻译成蛋白质。

(3)高等真核生物 DNA 所含的特定基因(如 rRNA、组蛋白基因等),在一定条件下会发生扩增,扩增可能与多基因家族的形成有关。

(4)假基因(pseudogene)是真核生物 DNA 的另一个特点。这些序列与真基因具高度的同源性,但由于突变而不能表达。有人估计,4 个 DNA 序列中有一个是假基因。假基因的存在扩充了真核基因组的大小。

以上讨论说明,真核生物的基因组,无论是其大小,所含的重复序列和单一序列都显示出与原核生物有很大不同。同时,高等真核生物基因组中真正编码蛋白质的序列仅占很小部分,估计不超过 6%~7%。高等真核生物基因组的这一特点曾引起不少人的注意。曾把它称为自私的 DNA(selfish DNA),即 DNA 似可极度扩展其自身至每一种生物可承受的极限。那些功能不清而与表型无关的 DNA 序列曾被称为废弃 DNA(junk DNA),但这部分 DNA 序列是否是毫无功能的呢? 这一问题,现在似乎正在逐步揭开面纱。

第八节 DNA 的精细结构

一、依赖于序列的 B-DNA 构象变化

Watson-Crick 的 DNA 双螺旋结构模型是根据 DNA 纤维的 X 射线衍射分析建立起来的,这样得到的是双螺旋结构的平均信息,从而归结出 DNA 结构的规整性。到了 70 年代末、化学合成技术的改进使人们能够合成具确定序列的 DNA 片段,并使之形成单晶。这种合成 DNA 晶体的 X 射线衍射揭示,DNA 双螺旋实际上不是完全均匀的,即双螺旋的许多结构参数是随碱基序列的不同而在一定范围内变化的,这称为 DNA 的局部构象。在 DNA 中,随碱基序列不同而

变化的参数有许多，其中重要的有螺旋扭转角(helix twist angle)，螺旋桨扭角(propeller twist angle)，碱基对转动角(roll angle of base pair)等，此外碱基对间还会发生滑动(图 4-20)。

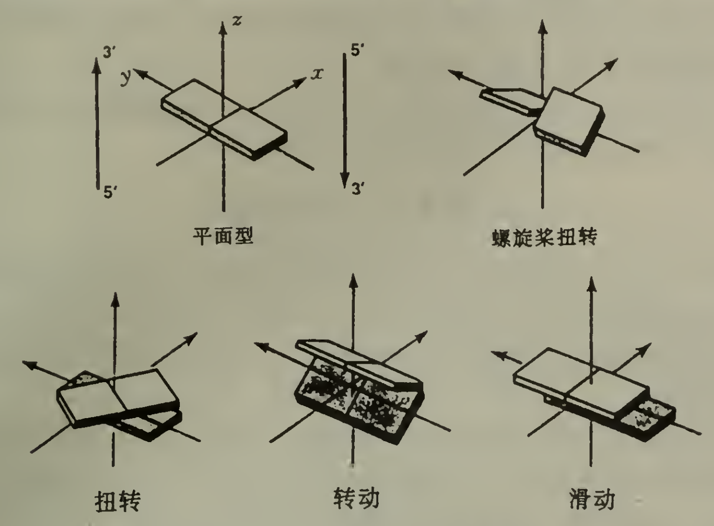


图 4-20　DNA 的一些局部构象变化

具 CGCGAATTCGCG 的 DNA 片段的 X 射线衍射分析指出，它的螺旋扭转角是在 28～42°间变动的(图 4-21)，而不是 36°，其它的参数也会随序列的不同而变动，只是数值不同。但 DNA 的局部构象为什么会显示出依赖于序列的变化呢?

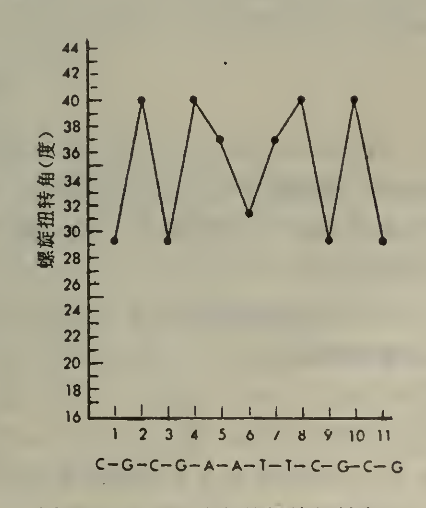


图 4-21　DNA 片段的螺旋扭转角

前面已讨论过，DNA 的双螺旋结构是通过碱基对的堆积和形成氢键而稳定的。但碱基的堆积会因序列的不同而有变化。这是由于碱基对中，嘌呤是双环结构，它大于单环的嘧啶，故而在碱基对中会伸过中点突出到对方一侧。这一情况不仅增加了碱基的堆积程度，也使碱基官能团间发生挤压，而这种挤压是随序列而变化的。对嘌呤-(3′,5′)-嘧啶序列言，来自鸟嘌呤 O_6/腺嘌呤 N_6 间的挤压发生于大沟中；对嘧啶-(3′,5′)-嘌呤序列言，来自鸟嘌呤 N_3,N_2/腺嘌呤 N_3 间的挤压发生于小沟中(图 4-22)。位于小沟的挤压更

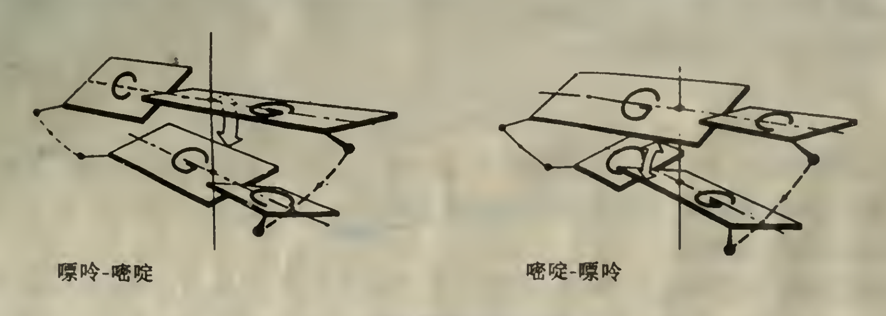


图 4-22　碱基官能团间的挤压

为严重。现已观察到，这种挤压发生于混合嘌呤/嘧啶序列中。

为了缓解这种效应，双螺旋结构就做出局部调整，以增强有利的碱基堆积，减少不利的挤压，如改变螺旋扭转角、碱基转动角、碱基对的螺旋桨扭角和相邻碱基对间发生滑动等。双螺旋中也可能除了标准碱基配对外，形成交叉氢键。

以上这些变化使DNA形成了精细结构(fine structure)。而精细结构正是DNA发挥功能时，相应蛋白质因子与靶位点作专一性识别和结合的标志。

二、连续AT序列的构象

现已观察到，DNA与一些蛋白质因子的结合部位是弯曲的。这种弯曲(bend)是连续A残基产生的。产生弯曲需要至少有3个连续的A(A_3)，最大的曲率是由连续的A_6造成的。产生弯曲的另一个要求是连续的A为其它的核苷酸所隔开。现观察到的最佳状况为5 bp，其它核苷酸所隔开的是连续A_5。

也有这样的情况，有些DNA序列本身并不弯曲，但在有关蛋白质的诱导下会发生弯曲。这一情况首先在核小体中观察到。为了实现组蛋白八聚体与DNA的结合，面向蛋白质的DNA沟变窄，而另一面则扩大。DNA沟的这种变化是和其所含的一些短序列有关。在DNA的小沟朝内的部位常含2～3个连续的AT碱基对，在大沟朝内处则常有2～3个连续的GC碱基对。核小体常位于这种变形区内。这说明，组蛋白不是随意结合于DNA的。

三、含错配碱基对的B-DNA

含GT，AC和GA碱基对的寡聚核苷酸的晶体分析说明，这些错配的碱基对都能包容在B-DNA双螺旋中，它们的构象参数一般处于允许范围内。每一错配碱基(除了一种GA配对外)都形成两个氢键，并且它们的宽度也都在Watson-Crick碱基配对宽度的0.05 nm范围内(图4-23)。

GA可采取两种配对方式，两种方式中G都为反式，而A则一个采取反式；另一个采取顺式；前者的宽度为1.25 nm，后者为1.07 nm。已知

5′CGC**G**AATT**A**GCG

3′GCG**C**TTAA**G**CGC

两个GA碱基对中A的苷键都采取顺式构象，而G则仍保持反式构象。这样，形成的GA碱基对仅宽1.07 nm，与GC对的尺寸十分接近，亦未引起螺旋方向的改变，所以，这一错配于DNA的结构并不产生明显的影响。在DNA的GA碱基对处，腺嘌呤官能团的位置与螺旋的其余部分明显不同，这一特点可能为修复酶提供了识别标志。GA碱基对容纳进螺旋的方式可通过下例说明。在

5′CCAA**G**A**T**TGG

3′GGTT**AG**AACC

例中，两个错配碱基对中的G和A都采取反式构象，它们通过拓宽小沟而容纳进螺旋。G和A间，除了形成两个氢键外，扩大的螺旋桨式扭角也使G的N_2和相邻AT中T的O_2间形成氢键，从而使螺旋结构稳定化。

错配碱基对包容于双螺旋，以及它们所产生的构象变化，明显地依赖于各自的序列环境(sequence environment)，这方面的知识对复制过程的校对、复制后的修复和重组的酶学都是

需要的。

图 4-23　错配碱基对间的氢键和量度

四、再论 Z-DNA

Z-DNA 的形成依赖于一定的序列。Z-DNA 与 B-DNA 的差别不仅在于螺旋的方向，亦涉及苷键的取向。在 A-DNA 和 B-DNA 中，苷键的构象都属反式，糖环属 C2′-内式。在 Z-DNA 中，嘧啶碱基的苷键仍为反式和糖环为 C2′-内式，而嘌呤核苷酸的相应构象则转换为顺式和 C3′-内式。由于相邻碱基对间的螺旋转角变化很大，因而形成一种二核苷酸重复单位，出现了螺旋转角的交替变化。在 CpG 中，C 和 G 间的螺旋转角为 15°；在 GpC 中，G 和 C 间的转角为 45°。交替的嘌呤-嘧啶序列虽然倾向于形成 Z-DNA，但也可由于邻近的 B-DNA 区域而被迫采取 B 型。此外，序列以外的因素，如甲基化、特定的 DNA-结合蛋白的存在和超螺旋程度的改变等都能影响 DNA 所采取的构象。

五、DNA 的局部构象与 DNA-结合蛋白

DNA 的局部构象是有其功能上的意义的，它们常是 DNA-结合蛋白（DNA-binding protein）结合的标志。DNA-结合蛋白泛指一类能结合于 DNA 的蛋白质，它包括各种酶（如 DNA 聚合酶、RNA 聚合酶等）和调节蛋白（如 CAP，Cro 阻遏物等）。这些蛋白质与 DNA 的结合涉及到生物学中的一些基本问题，如遗传和个体发育等，因此，核酸-蛋白质交互作用（nucleic acid-protein interaction）已成为分子生物学中的一个研究热点。DNA 和蛋白质的结合实质上是一个双向过程，DNA 的局部构象既提供了识别标志和发生交互作用的结构条件，而蛋白质的结合又促使该处 DNA 的构象发生进一步的变化和使交互作用深化。如大肠杆菌的调节蛋白 CAP（catabolite gene activator protein）对乳糖操纵子（lac operon）的启动子（promoter）的结合即导致该DNA 区段产生更多的折叠。核小体中 DNA 与组蛋白和限制性内切酶 *Eco*RI 与

DNA 上识别位点所形成的复合物(complex)的 X 射线结晶学分析就为这种相互关系提供了更为直接的证据。

限制性内切酶 *Eco*RI 是含两个相同亚基的二聚体,它结合于 DNA 的识别位点(recognition site)上。

TCGC | GAATTC | GCG
GCG | CTTAAG | CGCT

如有 Mg^{2+} 存在,酶就在特定位点切割 DNA,故 *Eco*RI-DNA 复合物是在无 Mg^{2+} 条件下分离得到的。衍射分析结果表明,*Eco*RI 的两个亚基紧紧包裹着 DNA 双螺旋,它与大沟中的碱基契合而与小沟无接触。在复合物中,DNA 的构象不同于经典的 A 型和 B 型;变形集中于三个突然出现的扭结中,后者又把各含 3 bp 的 4 组核苷酸分隔开。由于 DNA 序列的对称性,第一和第三扭结完全相同但不同于中间的扭结。DNA 的两个相同的末端片段具有类似于 A 型的构象,而中间的区域类似于 B-DNA。两个末端扭结的构象反映为 A 型和 B 型的过渡状态。结合时,位于 B-DNA 的中心结的双螺旋骤然松开 25°和形成 12°的弯曲,结果使大沟扩大以使 *Eco*RI 得以接近(图 4-24)。这一结构明显地不同于与限制酶复合前的情况。中心结的构象尤为奇特,它在未与 *Eco*RI 结合前可能仅十分短暂地存在。B-DNA-*Eco*RI 复合物的结构说明了蛋白质与 DNA 的交互作用决定于多方面的因素,其中包括 DNA 的特定序列和蛋白质的结构。结晶学分析则为 DNA 的扭结提供了更为直接的证据。

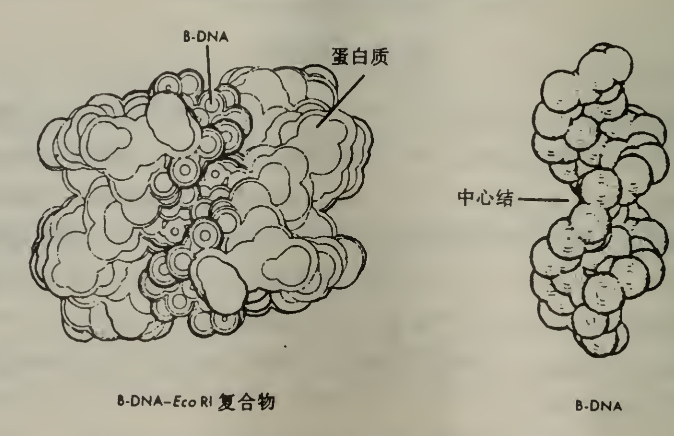


图 4-24 B-DNA-*Eco*RI 复合物中两者的交互作用

DNA 是遗传信息的载体,而遗传信息则是通过核苷酸的排列顺序体现出来的,DNA 所携带的遗传信息,根据已有证据大致可以分为两类:一种是编码蛋白质,它们通过核苷酸三联体与各个氨基酸间的对应关系使遗传信息由 DNA 流向蛋白质,但遗传信息的这种流动不是通过基因自身的功能,而是通过独立于该基因之外的蛋白质合成机构实现的。同时,蛋白质基因的表达并不总是构成型的,它是受调控的。DNA 所携带的第二类信息即是有关基因表达调控的信息,它们贮存在 DNA 的精细结构中,直接表现为特定的空间结构,即密码结构域(code domain),然后为相应的蛋白质因子所识别和结合,从而控制蛋白质基因的表达,这一方面的研究近来已取得不少进展,并已触及"废弃 DNA"是否具有一定的功能。无疑,这将成为分子生物学中的一个重要领域。

第九节 超螺旋 DNA

一、超螺旋 DNA 的发现

超螺旋 DNA(supercoiled DNA)最早是在 SV40 和多瘤病毒中发现的。人们使用电子显微镜可直接观察到超螺旋 DNA 的形象。

当 SV40 DNA 从被感染细胞分离出来时，它与组蛋白结合形成为核小体，每一分子约形成 24 个核小体。如使 DNA 与组蛋白分离，DNA 就成为超螺旋(图 4-25)。起初，人们认为超螺旋是一些小环形 DNA 的特点。以后发现它是几乎一切 DNA，无论是环型或线型 DNA 的共有的重要特征。

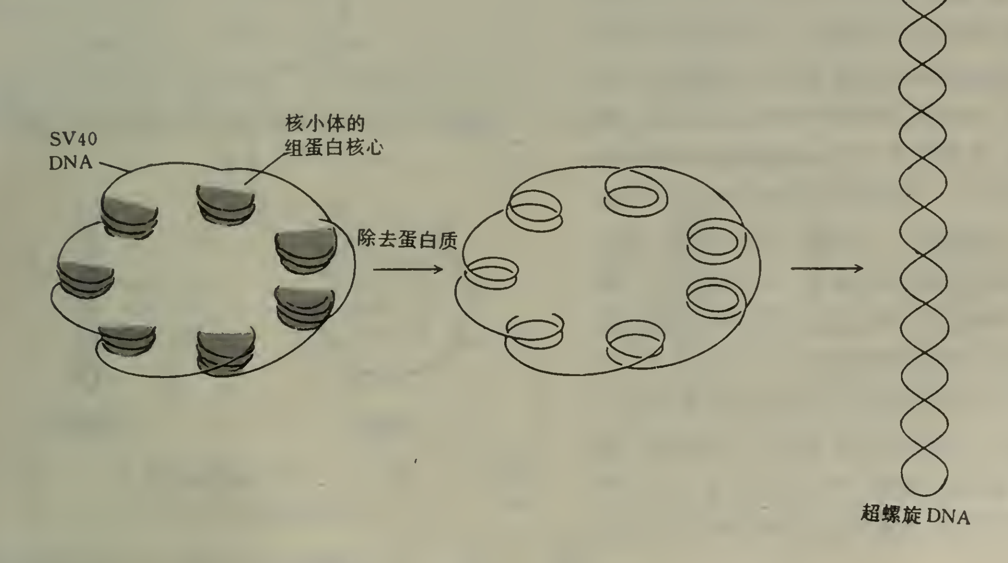


图 4-25 SV40 DNA 形态转变

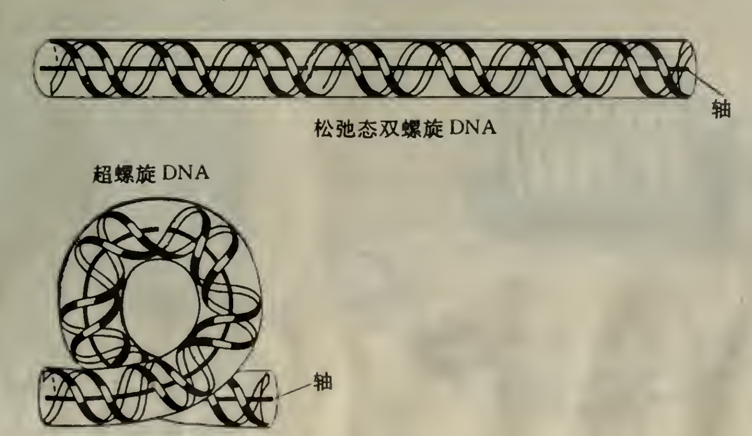


图 4-26 双螺旋和超螺旋 DNA

DNA 的分子是一种双股螺旋，但当它自身进一步盘绕时就形成超螺旋，DNA 双螺旋和超螺旋见图 4-26。

二、DNA 超螺旋的形成

B-DNA 的多核苷酸链的空间结构是以一组构象参数为表征的，它规定 B-DNA 的每匝螺旋含 10 bp(晶体中)，这是一种松弛状态。它的轴成一直线。对松弛状态的环型 DNA 言，轴处在同一平面中，故可以平放在平面上，但如 B-DNA 每匝的螺旋数通过一定的途径改变，它就可以形成超螺旋(图 4-27)。例如，一环型双股 DNA 共含 420 bp，它应形成 42 匝螺旋。现在，如切割其骨架中的任一磷酸二酯键，它就转变为含同数碱基对的线型 B-DNA。如使所形成的线型 DNA 的一端固定另一端向左放松 6 圈，然后再使两端重新闭合，由于 B-DNA 是一种在能力学上稳定的结构，螺旋数的减少就使其转变为一种受力状态。分子所经受的张力可以按两种方式分布：一种方式是分子保留一单链区，其余部分仍保持 B-DNA 状态；另一种方式是形成超螺旋，超螺旋的形成使已放松的双螺旋分子的碱基对接近于 B-DNA 中的状态(图 4-27)。

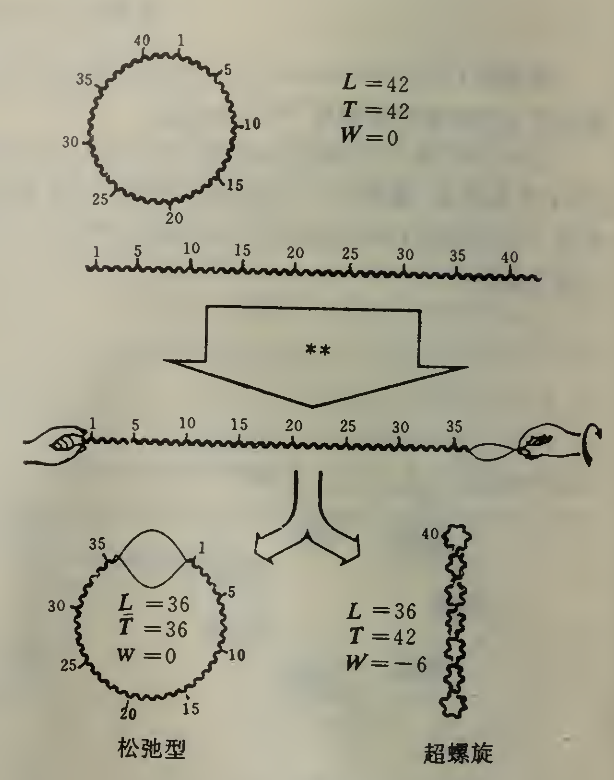


图 4-27　DNA 超螺旋的形成

* * 右手双螺旋放松 6 圈 $\Delta L=-6$。

超螺旋 DNA 可采取两种拓扑学上相当的形式：一种相当于双螺旋绕圆柱体旋转；另一种相当于双螺旋相互盘绕。超螺旋的这两种形式可以相互转变(图 4-28)。



图 4-28　超螺旋 DNA 的两种拓扑形式

DNA 结构的上述变化可以用数学式来表述。

$$L=T+W$$

L 称为 DNA 的连接数(linking number)，它是 DNA 的一股链绕另一股链盘绕的次数，它代表环型双螺旋 DNA 分子的拓扑特性。在不发生链的

断裂时，它是一常数。例如，图 4-27 中 DNA 原来的 L 为 42，放松后为 36。在右手双螺旋中，L 规定为正，T 称为盘绕数(twisting number)；它代表 DNA 的一股链绕双螺旋轴所做的完整的旋转数。对 B-DNA 言，它等于 DNA 的碱基数除以 10(晶体中)。W 为超盘绕数(writhing number)，代表双螺旋轴在空间的转动数，T 和 W 是可变的。

现在，我们回到图 4-27，其中原来的环型 DNA 的 $L=42$，$T=42$，故 $W=0$，即不存在超螺旋。它螺旋轴处在同一平面中，分子处在松弛状态。在保留一单链区的 DNA 分子中，L 和 T 都为 36，分子未形成超螺旋。在超螺旋 DNA 中，L 仍为 36，如使 T 保持未松开前的 42，W 就成 -6，即成为超盘绕 6 次的负超螺旋。

超螺旋是 DNA 三级结构的一种普遍形式。双螺旋 DNA 的松开导致形成负超螺旋，而 DNA 的拧紧则导致形成正超螺旋。

天然的 DNA 都呈负超螺旋，但在体外可得正超螺旋。溴乙锭、放线菌素 D 等的分子都是扁平的，它们可以嵌入 DNA 的碱基对之间。溴乙锭分子的嵌入能使相邻碱基对的间隔增至 0.7 nm，即使螺旋发生了解旋。对一个呈负超螺旋的环形 DNA 分子来说，溴乙锭的嵌入并没有改变连接数，但盘绕数减少，结果超螺旋数向相反方向改变。随着溴乙锭量的增加，负超螺旋 DNA 就转变为松弛态；溴乙锭的进一步增加，DNA 就转变为正超螺旋(图 4-29)。

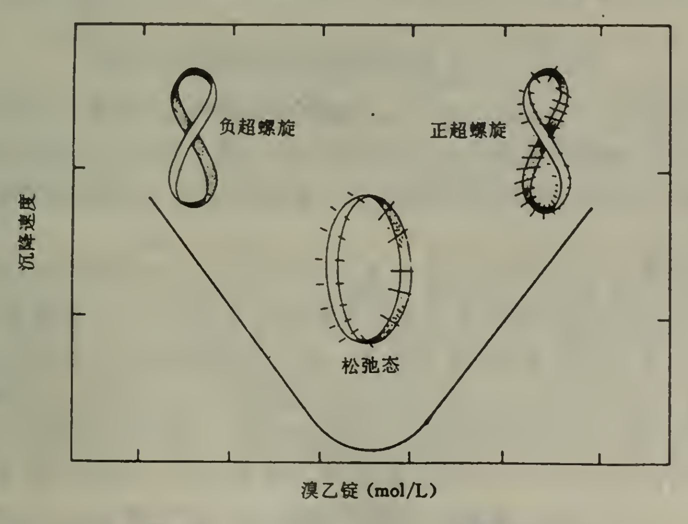


图 4-29 通过溴乙锭的嵌入，环形 DNA 超螺旋状态的改变



图 4-30 松弛的和超螺旋 DNA 的凝胶电泳

环形 DNA 分子会由超螺旋化而变得更为致密，它们在超离心中的沉降速度和在凝胶电泳中的迁移速度都增加，故超螺旋 DNA 可以通过这两种方法来检测和分离。现在，琼脂糖凝胶电泳已成为检测 DNA 超螺旋程度的一种最直接的方法(图 4-30)。它能够分离仅差一圈的超螺旋 DNA。

对一特定的 DNA 言，L 是一定值，但细胞内存在着一些可使 DNA 分子断裂和重新连接的酶，故 L 会在一定范围内变化，DNA 的超螺旋结构常处于动态变化中(图 4-31)，在这一过程中，

我们实际上关心的是连接数的改变及其对 T 和 W 的影响，即

$$\Delta L=\Delta T+\Delta W$$

不过，超螺旋对 DNA 的影响随其分子的长度而异。为了表达一定的超螺旋(W)对不同 DNA 分子的影响，人们就引入超螺旋密度(superhelix density)σ 概念，

$$\sigma=\frac{\Delta L}{L}$$

$\Delta L=L-L_{o}$　　L_{o} 为松弛态 DNA 的连接数

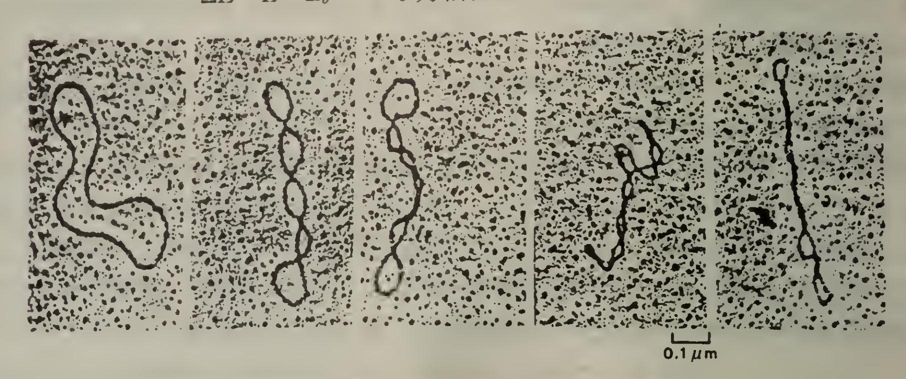


图 4-31　DNA 的电镜图(大肠杆菌的质粒二聚体)

当两个不同长度的 DNA 做比较时，L 代表螺旋和超螺旋的总数，它与分子的长度和其拓扑状态两者都有关系，故情况比较复杂。如采用 σ，由于它仅与分子的拓扑张力有关，故而可直接加以比较。现在，已测得从细胞和病毒粒子分离到的 DNA 的超螺旋密度介乎 0.05～0.07 间，一般为 0.05。

虽然超螺旋最早是在环型 DNA 中发现的，但其后的工作证明线型 DNA 也是超螺旋化的。对用温和方法分离得到的真核染色体的考察指出，它们的 DNA 沿着由 DNA-结合蛋白组成的支架(scaffold)形成许多连续的环状区域。每一个环中的 DNA 独立形成超螺旋和具有不同的张力，支架的存在可以防止一个环中的张力传递到另一个环中去。从果蝇的巨大唾液染色体和蝾螈的环状灯刷染色体所获得的许多证据说明，每一个环代表染色质的一个功能单位，常转录出一个或更多的长 RNA 分子。现已查明，染色质 DNA 中与蛋白质结合的区域(称为支架附着 DNA，或 SAR DNA，即 scaffold attached region DNA)是弯曲的，这表明 DNA 是在特定的区域与蛋白质结合的。

三、超螺旋的意义

对 DNA 言，超螺旋是普遍存在的，这也就表明了它的重要性。

在活体中，DNA 的结构不是一成不变的。DNA 的各种构象是互变的，是动态的。DNA 的这种结构变化对它的功能具有重要意义。而 DNA 超螺旋程度的改变介导了这种结构变化。B-DNA 是一种能力学上稳定的结构，超螺旋的引入就提高了它的能量水平。负超螺旋的存在会影响 DNA 结构变化的平衡，具超螺旋的 DNA 能实现松弛态 DNA 所不能实现的结构转化。

已如前述，大多数 DNA 的超螺旋密度约为－0.05，这相当于每 200 bp 存在 1 负超螺旋，其结果就赋予 DNA 分子以约 37 681 J/mol 能量。DNA 的超螺旋水平在活体中是重要的，但它并不是在整个 DNA 分子中都是均匀的。DNA 特定区域中超螺旋的增加有助于 DNA 的结

构转化。DNA 结构变化之一就是使DNA 双股链分开，或局部熔解。超螺旋所具有的多余的能量被用于碱基间氢键的断裂。DNA 中，10 bp 的分离大约需 50 241.6～209 340 J/mol，因此，DNA 所具有超螺旋水平仅是分离很少几个碱基对，但 DNA 的这种结构上的变化对复制、转录等的启动仍很重要。

DNA 结构变化中的一个极端例子是 B-DNA 向 Z-DNA 的转变，负超螺旋有利于形成 Z-DNA。B-DNA 转化成 Z-DNA 仅发生于特定的序列处，在连接处这种转化涉及到右手性双螺旋转变为左手性双螺旋，即 T 值改变 2。所需的自由能较高，约 32 238.4 J/mol，以后每延伸 1 bp 对仅需 1 884.1 J/mol。

超螺旋不仅使 DNA 形成高度致密的状态从而得以容纳于有限的空间中，在功能上也是重要的，它推动着结构的转化以满足功能上的需要。

第十节　拓扑异构酶

一、拓扑异构酶的分类

细胞内存在着一类能催化 DNA 拓扑异构体(topoisomer)相互转化的酶，它们称为拓扑异构酶(topoisomerase)。它们能与 DNA 形成共价结合的蛋白质-DNA 中间体，从而在其骨架的磷酸二酯键处造成暂时性的裂口，使 DNA 的多核苷酸链得以穿越，结果改变了分子的拓扑状态。在这一过程中，DNA 的连接数虽然改变了，但其核苷酸序列并无任何变化。拓扑异构酶还能使 DNA 发生连环化(catenate)或脱连环化(decatenate)，打结(knot)或解结(unknot)。

拓扑异构酶可以分为Ⅰ型和Ⅱ型两类(表 4-4)。Ⅰ型和Ⅱ型拓扑异构酶的作用不同。Ⅰ型酶在 DNA 的一股链产生一切口，从而使另一股链得以穿越，而Ⅱ型酶则在两股链上都产生一切口；其结果是，Ⅰ型酶使 DNA 的连接数每作用一次改变 1，Ⅱ型酶则使连接数每次改变 2。在细胞中，这两类酶的活力是严格调节的，拓扑导构酶Ⅱ使 DNA 超螺旋化的作用为拓扑异构酶Ⅰ使 DNA 松弛化的作用所抗衡，从而使 DNA 的超螺旋程度保持在适当的水平上。由于 DNA 的超螺旋程度深刻地影响着机体的活动，故能改变拓扑异构酶活力的突变常是致命的。

表 4-4　拓扑异构酶的类别和特性

性　质	Ⅰ型		Ⅱ型	
	原核生物[1]	真核生物[2]	原核生物	真核生物
被切割的 DNA 链	1	1	2	2
亚基分子量(kD)	100	95	97，90	150
亚基数	单体	单体	A_2B_2	均二聚体
对 ATP 的需求	否	否	是	是
对 Mg^{2+} 的需求	是	否	是	是
依赖于 DNA 的 ATP 酶	否	否	是	是
产生负超螺旋	否	否	是	否
松弛负超螺旋	是	是	否[3]	是
松弛正超螺旋	否	是	是[4]	是
连环、打结	是[5]	是[5]	是	是

1. 大肠杆菌 topo Ⅰ 和 topo Ⅲ；2. 酵母 topo 3 很可能编码特性更近似大肠杆菌的Ⅰ型酶；3. 无 ATP 时能松弛负超螺旋；4. 借助于引入负超螺旋；5. 要求双股链之一具有一切口。

二、原核拓扑异构物

最先发现的拓扑异构酶是大肠杆菌的 ω 蛋白，它能使超螺旋 DNA 松弛。以后在其它细菌和真核细胞中都发现了能使 DNA 松弛的酶，故 ω 蛋白就重新命名为大肠杆菌拓扑异构酶 Ⅰ（topo Ⅰ），由于大肠杆菌中能使 DNA 超螺旋化的旋转酶（gyrase）有时称为大肠杆菌拓扑异构酶 Ⅱ，故而另一个 Ⅰ 型酶就称为大肠杆菌拓扑异构酶 Ⅲ。

拓扑异构酶的作用是催化 DNA 骨架中磷酸二酯键的断裂和重新连接。与其它的酶不同，它们的作用并不导致新类型共价键的形成，而是在多核苷酸键中产生一暂时性的断口。在这一过程中，多核苷酸链不产生游离末端，而是酶本身共价结合于两末端之一（究竟连接于 3′ 或 5′ 端则依赖于特定的酶）。这时拓扑异构酶的一个酪氨酸残基即连接于切割产生的多核苷酸链末端的磷酸基上，使原来贮存于磷酸二酯键中的自由能得以保存，并转而用于断口的重新连接。

（一）大肠杆菌拓扑异构酶 Ⅰ

拓扑异构酶 Ⅰ 是分子量约 100 kD 的单一多肽链，含 3～4 个为其活力所需的锌原子。

拓扑异构酶 Ⅰ 的松弛活力具有如下特点：它消除 DNA 的负超螺旋而不引起 DNA 发生其它改变；DNA 的松弛是逐步进行的，磷酸二酯键的自由能通过形成共价的蛋白-DNA 中间体而保存并转而用于链的重新连接，它不需要 ATP 或 NAD 这样的辅助因子。

拓扑异构酶 Ⅰ 的松弛作用机制大致如图 4-32。

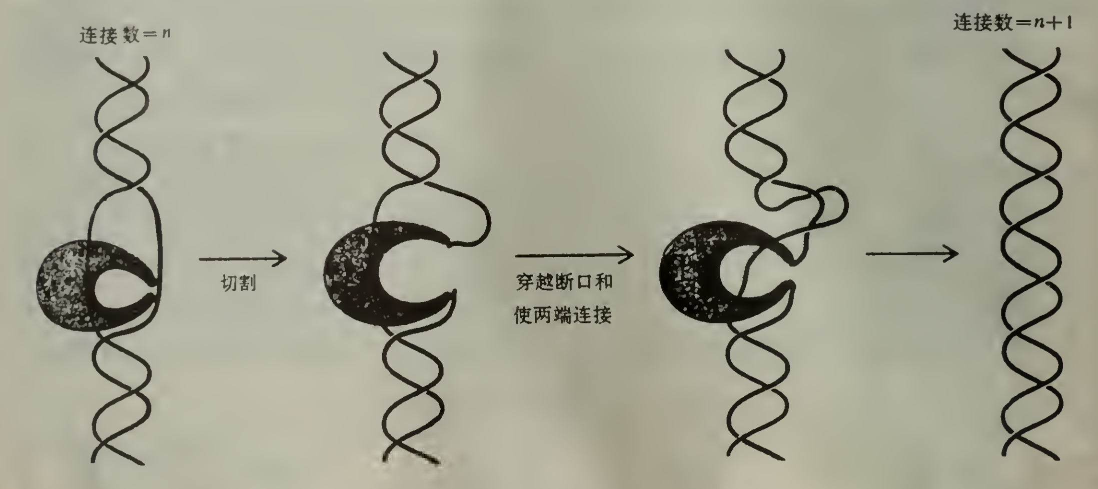


图 4-32　Ⅰ 型拓扑异构酶作用机理示意图

拓扑异构酶 Ⅰ 首先结合于 DNA，使该处的 DNA 熔解，随后与单链区形成酶-DNA 复合物；切割 DNA 双股中的一股，切割点的 5′-P 形成磷酰酪氨酸键；DNA 的一股链穿越切割点；断裂的链重新连接，酶被释放。

拓扑异构酶 Ⅲ 是大肠杆菌中另一个 Ⅰ 型酶，单体质量为 74 kD，但它似乎并不是必要的。

（二）大肠杆菌的旋转酶

旋转酶有时也称为大肠杆菌拓扑异构酶 Ⅱ，它使松弛的双股环型 DNA 转化为负超螺旋 DNA，故是 Ⅱ 型酶。它的亚基结构为 A_2B_2，质量分别为 97 和 90 kD。亚基 A 的基因 *gyr A* 控制对萘啶酮酸（nalidixic acid）的抗性，故过去曾用 *nalA* 表示；亚基 B 的基因 *gyrB* 由于与香豆

霉素(coumermycin)的关系而曾表示为cou。

旋转酶与拓扑异构酶Ⅰ不同,拓扑异构酶Ⅰ使DNA的连接数每次改变1,而旋转酶则每次改变2(图4-33)。

旋转酶使DNA超螺旋化的步骤为:(1)旋转酶结合于DNA,并使长约105～140 bp的片段按正方向包裹自身;(2)在包裹片段的中心附近切割,每一断口的5′端与亚基A的Tyr 122共价结合;(3)接近或位于包裹片段内的DNA区域,通过亚基B与ATP的结合而穿越断口;(4)被切开的DNA链利用蛋白-DNA复合物所贮存的能量重新闭合;(5)DNA的超螺旋化依赖于ATP的水解。旋转酶的亚基B具ATP酶活力。在无ATP存在时,旋转酶能使负超螺旋DNA松弛。

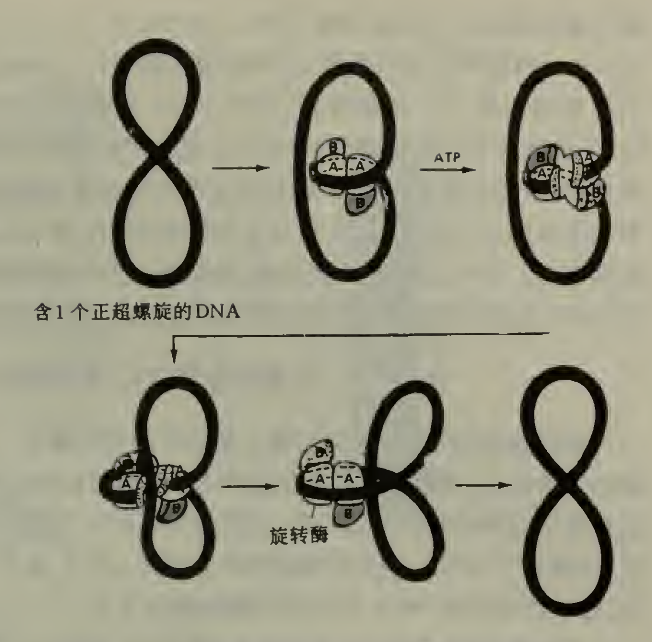


图4-33　由旋转酶催化的DNA超螺旋化示意图

为了清晰起见,起始DNA用1个正超螺旋代表。

三、真核生物拓扑异构酶

酵母、两栖动物、昆虫、哺乳动物和植物的Ⅰ型拓扑异构酶的特性相似,但它与原核生物的酶有明显差异。

真核生物拓扑异构酶Ⅰ是质量约为95 kD的单体蛋白。虽然它所催化的反应一般与原核生物的酶相似,但它同样能使正、负超螺旋DNA松弛。松弛作用不依赖于ATP,能发生于有EDTA存在的条件下,Mg^{2+}能提高酶的活力。酶通过其特定Tyr与断口的3′P连接而形成中间体,这与原核生物酶形成于5′P端不同。

Ⅱ型真核生物拓扑异构酶为150～180 kD的均二聚体,能以同样的速率松弛正和负超螺旋DNA;但与原核生物酶不同,不能产生负超螺旋。作用时,它需要ATP和Mg^{2+}。

原核和真核生物拓扑异构酶都参与DNA复制、转录和重组等。

第十一节　染色质的结构

一、染 色 质

染色体原来是指真核细胞染色后存在于核内的着色物体。它只存在于分裂中的真核细胞内。在非分裂细胞内,它是无定形的,并弥散于核内,只是在有丝分裂前才形成确定的形状。现在,染色体已被普遍用于指存在于病毒、细菌、真核细胞及其细胞器内所含的核酸分子。

DNA属生物大分子。例如,λ噬菌体的DNA含48 502 bp,同时DNA与病毒粒子的长度相差悬殊,为17.5∶0.19(μm)。对原核生物的大肠杆菌来说,DNA的上述特点更为明显,大

肠杆菌的基因组为双股环型 DNA，它约含 4.7×10^6 bp，周长约 1.7 mm，但大肠杆菌的长度约 2 μm，即前者约为后者的 850 倍，远比病毒大。真核细胞的 DNA 含量又比原核高。就最简单的真核生物之一的酵母言，它的 DNA 含量为大肠杆菌的 4 倍，果蝇和人的每一细胞的 DNA 含量约为大肠杆菌的 25 和 600 倍。许多植物和两栖动物的 DNA 含量更高，还有一点应指出的是，它们都是线型分子，即长度更大。但真核细胞的半径一般在 10～30 μm 间。以上数字充分说明，不论是原核还是真核生物的 DNA 都不是直接以环型或线型分子的形态存在于细胞中的。DNA 必须经高度压缩，即形成一定的高级结构后才能装入类核和核内。过去讨论过的 DNA 超螺旋所提供的压缩倍数显然还远远不够，超螺旋 DNA 还进一步形成更高级的结构。

二、核小体是 DNA 在细胞内存在的形态

虽然游离的 DNA 在能力学上趋向于形成松弛态，但不能仅因此而认为天然的 DNA 是松弛的。事实上，所有原核和真核细胞 DNA 都是以负超螺旋形态存在的。第九节已指出，DNA 超螺旋可以两种拓扑形式存在。在溶液中，DNA 主要以相互盘绕的形态存在，这种形态较为稳定；在细胞中，DNA 主要以螺旋管形态存在，DNA 通过与蛋白质结合而稳定化，在这两种形态中，螺旋管形式对 DNA 的压缩所起的作用更大。

在染色质中，与 DNA 相结合的蛋白质为组蛋白。组蛋白是一类较小的、荷正电的、富含精氨酸和赖氨酸的蛋白质。组蛋白是已知的最为保守的蛋白质中的一种。组蛋白可以分为 5 种基本类型，分别称为 H1，H2A，H2B，H3 和 H4（表 4-5）。

表 4-5 组 蛋 白

类别	分子量(kD)	氨基酸残基数	碱性氨基酸含量(%)	
			Lys	Arg
H1*	21.13	223	29.5	1.3
H2A*	13.96	129	10.9	9.3
H2B*	13.774	125	16.0	6.4
H3	15.273	135	9.6	13.3
H4	11.236	102	10.8	13.7

“*”标出的组蛋白的大小，在种间稍有差异，表中给出的是牛的组蛋白。

染色质在电子显微镜下观察时呈现为由 10 nm 的球状颗粒和 DNA 纤维组成的念珠状外观。这些球状颗粒称为核小体（nucleosome），核小体间的 DNA 称为接头 DNA（linker DNA）。核小体所缔合的 DNA 约为 200 bp（对不同的机体和组织言，DNA 的量约在 150～250 bp 间变动）。由于接头 DNA 易于为核酸酶所作用，因此，当染色质以微球菌核酸酶（micrococcal nuclease）等轻微处理后，染色质就产生一系列依次相差 200 bp 的长度不等的 DNA 片段。核酸酶的更广泛处理就产生高度被保护的长 146 bp 的 DNA 片段。这一 DNA 片段及其所结合的组蛋白称为核小体核心颗粒（nucleosome core particle），每一核小体分别含 2 个 H2A，H2B，H3 和 H4，称为组蛋白八聚体（histone octamer）和一个当接头 DNA 降解后即失去的 H1（图 4-34）。

X 射线衍射分析证明，DNA 的含 146 bp 片段以左手性螺旋管的形态包裹于组蛋白八聚体的外面（图 4-34）。对这一结构的仔细观察可以解释，虽然真核细胞不含能产生负超螺旋的酶，但它仍然以负超螺旋的形态存在。这是由于，当双螺旋DNA包裹于组蛋白八聚体外时，

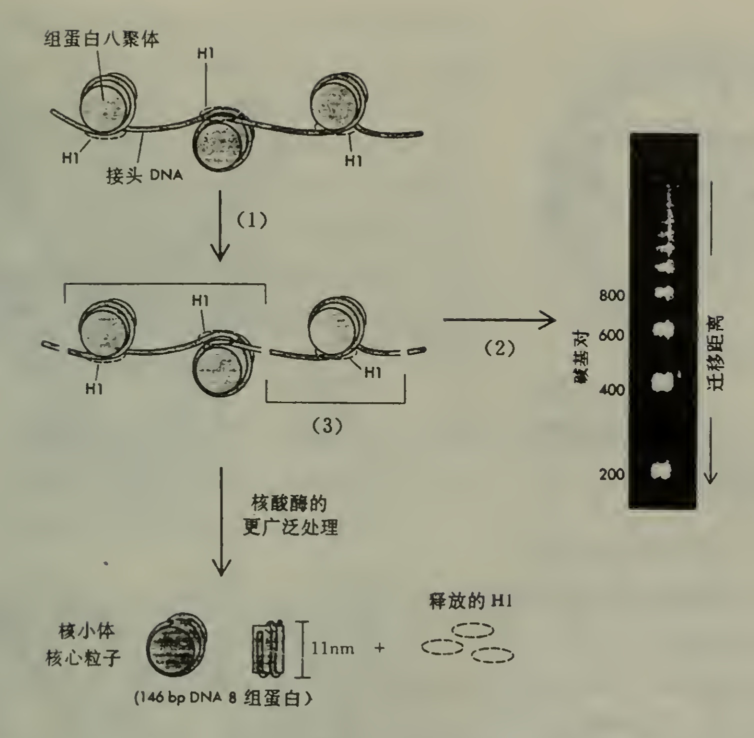


图 4-34 由 DNA 和核小体组成的念珠状结构

(1)核酸酶的轻微处理;(2)凝胶电泳;(3)切口间距离。

DNA 必须去除,匝螺旋才能完成。在体外,组蛋白核心与松弛态 DNA 的结合会造成一负超螺旋,但这一过程显然不需要切割 DNA 链,而是以另一处正超螺旋的形成为补偿的(图 4-35)。与原核 DNA 旋转酶不同,真核生物酶并不能使 DNA 负超螺旋化,但能使正超螺旋松弛,见表 4-4。现已在实验室中证明,环型 DNA 与组蛋白结合成染色质时,拓扑异构酶是必要的。拓扑异构酶使核小体形成时所产生的正超螺旋松弛,使整个分子的连接数降低,即形成负超螺旋。

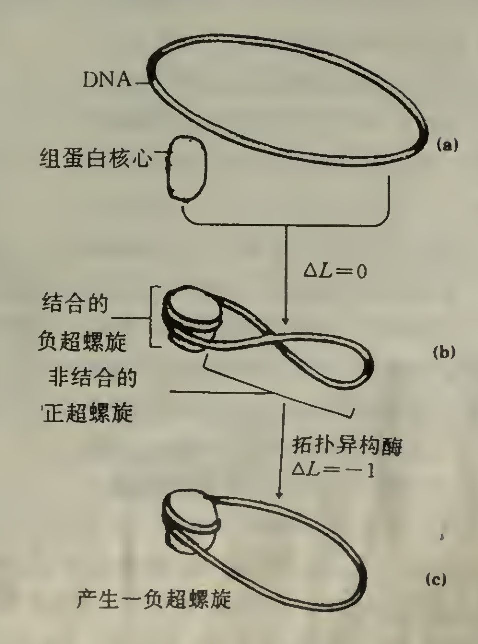


图 4-35 环型 DNA 与组蛋白结合成核小体

DNA 与组蛋白结合成核小体的另一个重要因素是涉及 DNA 的一定序列。组蛋白核心并不是随机地结合于 DNA 的各部分,而是专一地结合于特定位点。DNA 上核小体形成部位现在还不完全清楚,但有证据表明核小体形成于 DNA 中富含 AT 碱基对处,在该处 DNA 的小沟与组蛋白核心相接触(图 4-36)。DNA 紧紧包裹组蛋白核心就使该处的小沟受到压缩,显然 AT 碱基对有利于这种压缩的出现。

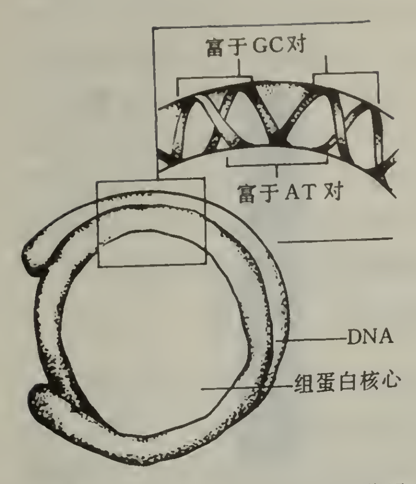
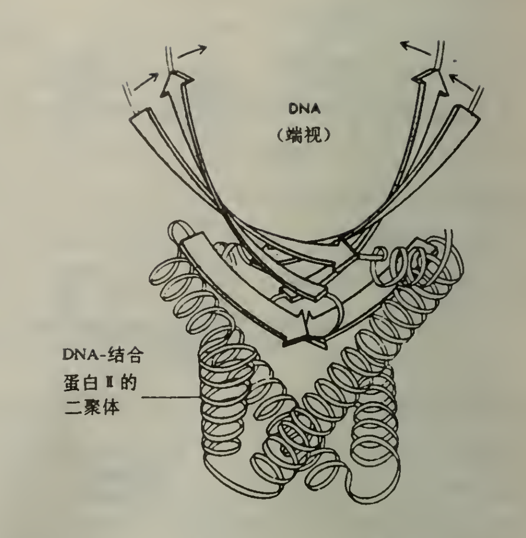


图 4-36　DNA 上形成核小体的位置



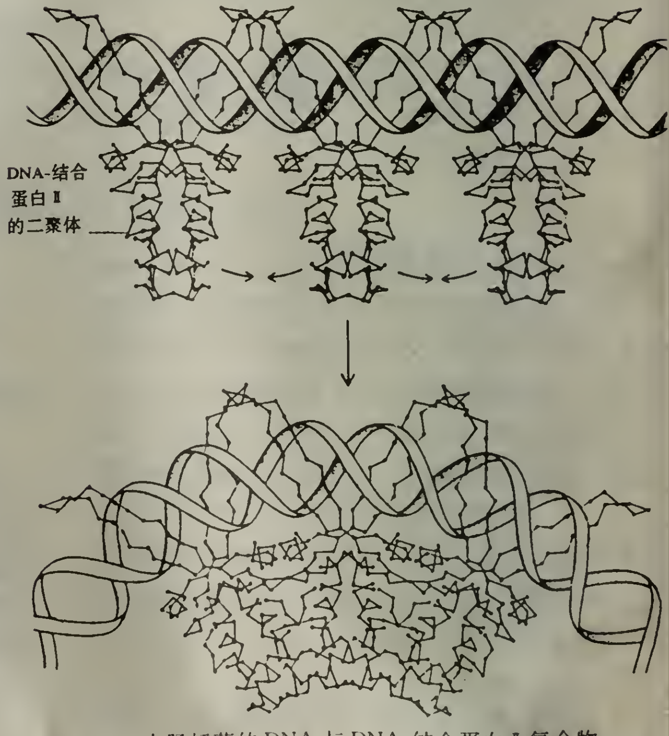


图 4-37　大肠杆菌的 DNA 与 DNA-结合蛋白Ⅱ复合物

在很长一段时间内，人们一直认为细菌 DNA 与真核 DNA 不同，它是裸露的，并不束紧成染色质样的结构。近年来，由于采用了更为温和的方法，人们已在大肠杆菌中观察到了形成珠状小束的凝集状染色体，并由此提取到了组蛋白样的碱性蛋白。细菌 DNA 与组蛋白样蛋白质的复合无可置疑地说明，DNA 为了发挥功能就要规整地束紧。从这些蛋白质中已分离得到一种称为 DNA-结合蛋白Ⅱ(也称为 HU 蛋白)的晶体。它是一种二聚体，每一亚基的分子量为 9.5 kD，它们各有一条延伸出的含精氨酸的臂。精氨酸在生理条件下带正电荷，DNA 主链的磷酸基带负电荷，因此，DNA-结合蛋白Ⅱ就通过静电交互作用与 DNA 双螺旋相缔合。关于 DNA-结合蛋白Ⅱ和 DNA 在复合物中的取向，现在认为大致是：蛋白质二聚体毗连地分布于 DNA 的一侧，蛋白质的另一端则通过静电作用而相互靠拢，结果就使 DNA 双螺旋发生弯曲而超螺旋化。这样的结合就使蛋白质位于复合体的内部(图 4-37)。有趣的是，所有天然存在的超螺旋 DNA，不论其来源，它们的平均超螺旋度(即每 10 bp 所形成的超螺旋圈数)都约为 0.05。这可能是由于原核和真核细胞中染色体的结构是十分相似的缘故。

还应该一提的是，病毒染色体不论在细胞中增殖或包装进病毒粒子时，也是和蛋白质复合的。SV40 的病毒染色体任何时候都束紧在 24 个含组蛋白的核小体中(图 4-38)，它们的结构与宿主细胞中的一样。当与蛋白质分开时，SV40 DNA 就呈现为超螺旋。相反，存在于腺病毒粒子中的 DNA 并未与组蛋白复合。这些脊椎动物病毒 DNA，通过所编码的一种独特的 DNA-结合蛋白而包装进荚膜中。可是，当腺病毒 DNA 在细胞中发挥作用时，它就释放出自身的 DNA-结合蛋白，并与各个细胞都相同的组蛋白结合。

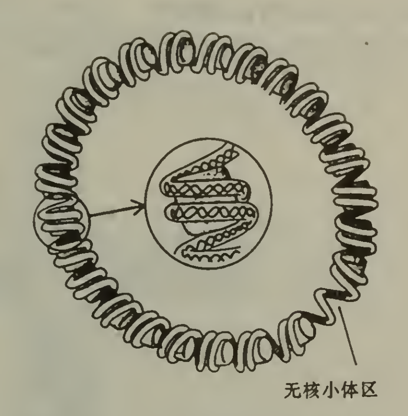


图 4-38　SV40 的含组蛋白的核小体
图中的无核小体区位于复制原点附近。

形成核小体是染色体结构的一个层次，它还形成更高级的结构(见第十三章)。染色体的这种高级结构不仅有利于 DNA 装入核中，并且还有着功能上的意义。

第十二节　一些 DNA 序列的不寻常结构

一、DNA 二级结构的多样性

DNA 的二级结构一度认为是均匀和单调的，可是现在已清楚，情况并非如此，DNA 的二级结构常是各具个性的。关于 DNA 的局部结构，第八节已有论述，本节着重讨论 DNA 的一些不寻常结构(unusual structure)。

现在已证明，DNA 的有些序列会形成特定的结构，这些不寻常结构不是完全独立于 B-DNA 之外的，有些以 B-DNA 为基础，有些由 B-DNA 转变而来。现已发现的 DNA 不寻常结构，如前已讨论过的 Z-DNA，十字形结构，单链臌泡(single stranded bubble)弯曲外，还有三链 DNA、四链 DNA 等。

DNA 的不寻常结构是以一定的 DNA 序列为基础的。例如，交替的嘧啶、嘌呤重复序列倾向形成 Z-DNA，反向重复序列倾向形成十字形结构，构成镜像重复的同型嘧啶-同型嘌呤序列可能形成三链结构，而富含 G 的序列可能形成四链结构等。这些结构的存在对 DNA 发挥其功能有着重要意义。也有些不寻常结构是在生命过程中形成于分子间的。例如，当二条具互补序列的 RNA 与双链 DNA 的一股链配对；另一股链就被游离成单链环，从而形成 R-环。转录过程中会出现 R-环。D-环是由双链 DNA 与另一 DNA 链形成的。D-环形成于哺乳类线粒体 DNA 的复制中，也是遗传重组中由大肠杆菌的 RecA 蛋白等所催化形成的。Holliday 结构是遗传重组中的重要中间体，它由 4 条 DNA 链组成(图 4-39)。两个 DNA 双螺旋的碱基可以通过连接部而配对并堆积。以上的分子间的不寻常结构是藉 DNA 间的同源序列形成的。

DNA 的不寻常结构的能量水平常高于 B-DNA，它的形成常藉 DNA 的负超螺旋来推动。如 DNA 的单链臌泡就是负超螺旋 DNA 张力的一种分布形式。DNA 分子中最先发生变性的常是含较多 AT 的区域，如这些片段含反向重复序列，则这些区段的变性就可能形成十字形结构(图 4-40)。十字形结构的碱基，除环中的少数几个外都是完全配对的。因此，十字形结构的形成过程中，氢键的减少并不多，超螺旋度稍有减少即足以提供其所需的能量。但这不足以说

明十字形结构在细胞中一定存在，这是一个还有争论的问题。因为这一转变可能在能力学上是可行的，在动力学上却是禁忌(kinetically forbidden)的。

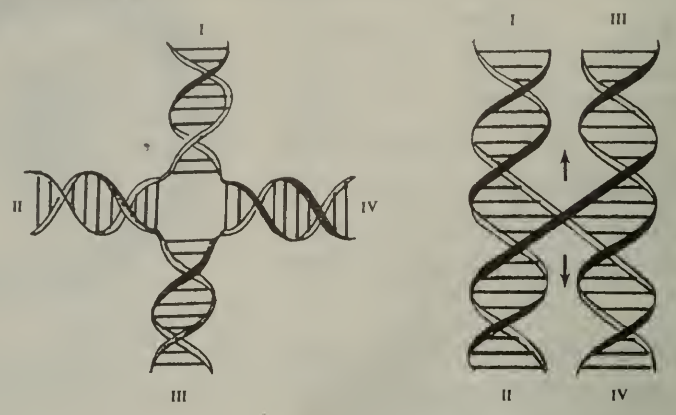


图 4-39 Holliday 结构

图的左方表示为平面构象，右方用两个平行螺旋表示，图中的箭头表示分支迁移。

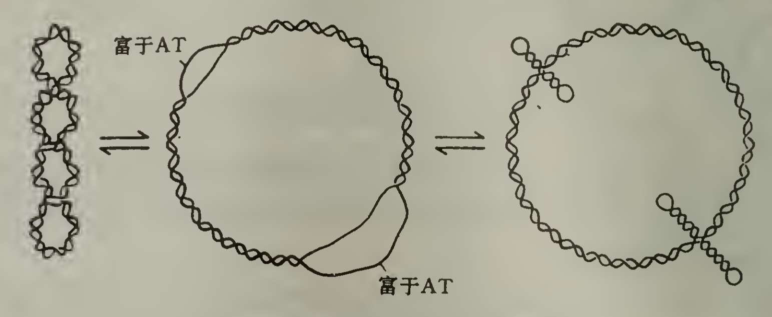


图 4-40 DNA 分子中十字形结构的形成

Z-DNA 的形成也常利用超螺旋的高能态。Z-DNA 是含 12 bp 重复的左手性双螺旋结构。过去认为，重复的CG 或 TA 对形成 Z-DNA 是必要的。现在发现这一条件并不是严格的，高达 8 个的 TA 如以 CG 重复为两翼时，也能形成左手性螺旋，只是比纯 CG 序列需要更多的能量。适当的序列在超螺旋DNA 张力的诱导下也能形成 Z-DNA。Z-DNA 是左手性螺旋，而 B-DNA 是右手性螺旋，B→Z 的转变降低了分子的螺旋匝数(T)。如 1.2 匝的右手性螺旋(12 bp)转变为 1 匝左手螺旋，T 值约改变 2。由于 L 保持不变，故 W 就升高 2，即负超螺旋减少 2。B→Z 的转变是协同性过程，B-DNA 和 Z-DNA 间的连接部构成了这一过程的能障(energy barrier)。Z-DNA 片段的稳定性是随其长度的增加而提高的。B-DNA 向 Z-DNA 的转变需要有足够的能量来克服连接部构成的壁垒。

许多不寻常结构的形成与 DNA 超螺旋值之间存在着类似的关系，这种关系可能与基因表达的调控有关。

二、三 链 DNA

早在 1957 年，人们就发现一股含嘌呤；另一股含嘧啶的双股多核苷酸，在体外能够形成三链 DNA。可是由于当时未能获得它可能存在于体内的证据，故这一发现沉寂了 30 年。1987 年，Mirkin 发现，一种质粒DNA 在酸性溶液中能形成三螺旋DNA。从此以后，这一领域的

研究才逐渐受到人们的重视。人们不仅发现在接近生理条件下能形成三链DNA，并且也发现在真核生物DNA中存在能形成三链DNA所需的同型嘌呤-同型嘧啶序列，其含量可高达1%。此外，也发现了其它的证据。三链DNA中的第三股链可来自分子内，也可来自分子外。

含$(TC)_n$和$(AG)_n$这样的同型嘧啶和同型嘌呤，并形成镜像重复(mirror repeat)的序列(图4-41)。在低pH条件下能形成分子内的三链DNA，

镜像重复

T T A G C A C G T G C T A A
A A T C G T G C A C G A T T

反向重复

T T A G C A C C A C G A T T
A A T C G T G G T G C T A A

图4-41　镜像重复和反向重复

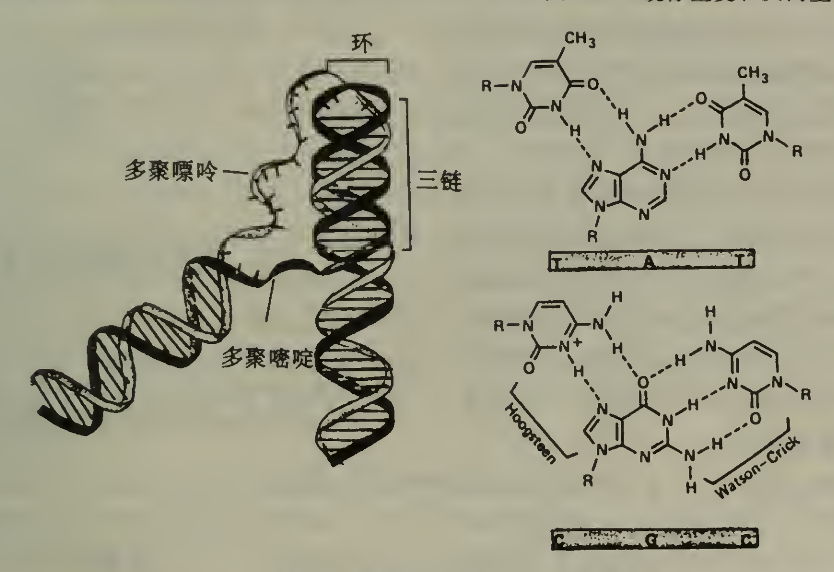


图4-42　三链DNA的形成和Hoogsteen碱基配对

它是由双链DNA拆开后产生的多聚嘧啶链回折并嵌入剩下的双链DNA的大沟中形成的(图4-42)。在三链DNA中，原来两股链的走向是反平行的，其碱基通过Watson-Crick方式配对，位于大沟中的多聚嘧啶链则与双链DNA中的多聚嘌呤链成平行走向，碱基则按Hoogsteen方式配对并形成TAT，CGC三联体。在后一配对方式中，多聚嘧啶链中的胞嘧啶残基必须先与H^+结合(质子化)才能与鸟嘌呤配对。这也就是H-DNA命名的由来。在形成三链DNA过程中游离出来的多聚嘌呤链则保持单链状态。故三链DNA对S1核酸酶的作用敏感。由于三链DNA含有镜像重复，故可以形成两种异构体：一是多聚嘌呤链的5′部分成单链；另一是3′部分成单链(图4-43)。

三链DNA的形成也可由负超螺旋的改变驱

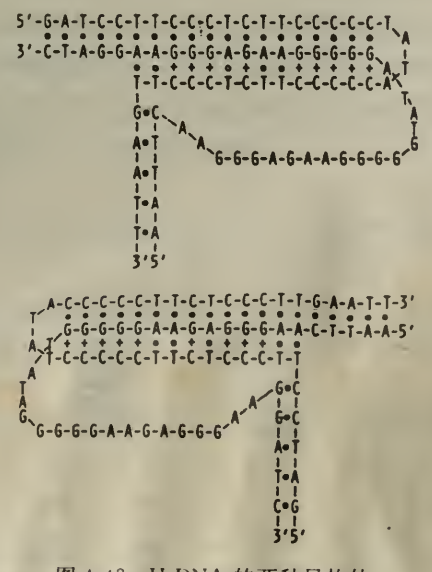


图4-43　H-DNA的两种异构体

动，随着负超螺旋的增加，即使在中性条件下 DNA 中 S1 核酸酶的超敏位点也能形成。

现在，已经在基因的调节区和染色体重组热点分离到能在离体条件下形成三链 DNA 的序列。这提示它们在基因表达中的可能作用，当然，这还待进一步的直接证据予以证实。

三、端粒 DNA 的结构——四链 DNA

端粒(telomere)是真核染色体的末端序列。端粒的生物学功能是保持染色体的稳定，它决定着细胞的寿命。线型 DNA 复制长期来存在的一个问题就是它的末端是如何复制的。DNA 的复制需要 RNA 引物，而引物随后被降解，其结果就使子代 DNA 的 5′端逐步缩短，端粒 DNA 的结构及其复制的阐明为解决这一问题奠定了基础。

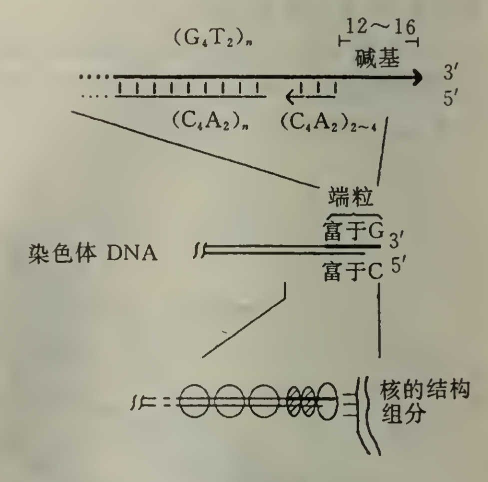


图 4-44　真核染色体的端粒

图下部表示染色体末端的端粒蛋白(用空白和斜线标出的椭圆形)和核小体(空白圆圈)。

与生物的多样性不同，端粒 DNA 序列和结构却是类似的。端粒 DNA 主要是由一段十分简单和串联重复的序列组成。例如，人和其它脊椎动物等的端粒重复单位为 AGGGTT，纤毛原生动物四膜虫的重复单位为 GGGGTT 和 GGGGTTTT，面包酵母等的重复单位为 $G_{1\sim3}T$ 和 $G_{1\sim8}A$。它们的共同特点是富含 G，它们的长度可达几百到几千碱基对。

端粒 DNA 的序列具有一定的取向特征。在每一染色体末端，富含 G 的一股链由 5′向 3′-末端延伸，并突出于互补的富含 C 一股链 12～16 核苷酸(图 4-44)。染色体的末端与特定的蛋白质形成复合物。

现已证明，端粒或类似端粒的寡聚核苷酸在实验室条件下能形成多种 G—G 碱基配对形式。例如，短的如端粒 DNA3′-末端那样的 12～16 bp 突出序列能够形成回折结构(图 4-45 b)，它们可能是通过碱基间的非标准配对形成的。较长的富含 G 的序列可能形成 G 的四链 DNA(图 4-45 c，d)。在这样的结构中，G 形成一种四联体，相互间通过 Hoogsteen 方式结合，并为

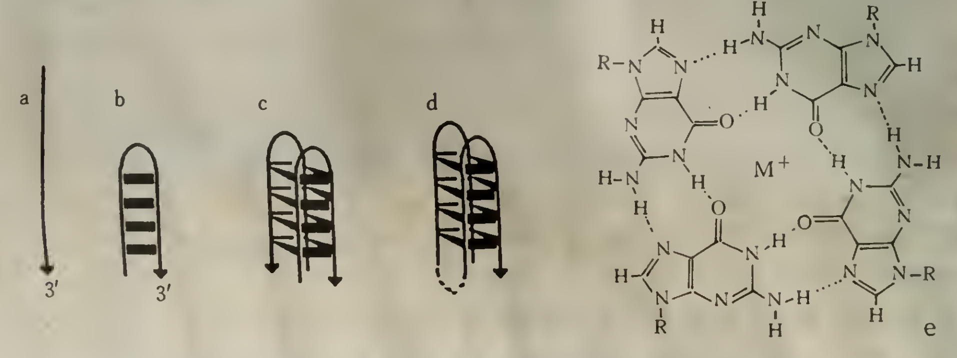


图 4-45　富含 G 的端粒序列形成的不寻常结构

a 为单链形式；b 为回折形成的结构；c，d 为所形成的四链结构；e，G-四联体。

K^+或Na^+所稳定(图 4-46),四联体再上下堆积。

端粒DNA 究竟具有何种结构有待更深入地研究才能确定。关于端粒DNA 的复制则将在第五章中讨论。

四、DNA 序列与遗传信息

DNA 是遗传信息的载体。但它载有哪些遗传信息和又怎样荷载这些遗传信息呢? 这确是一个值得而又必须提出的问题。

长时期来,人们认为DNA 所携带的遗传信息主要是通过编码蛋白质体现的。但编码蛋白质的序列仅占真核DNA 的很小一部分。这是否表明DNA 的其它序列都是无用的呢? 两方面的证据终于使这一问题的答案露出了端倪。一是原核生物基因表达调控中阻遏物是和DNA中的特定序列(操纵基因)结合的,这种结合具有特异性;二是DNA 具有依赖于序列的精细结构(见第八节)。这说明DNA 的一些序列可能与基因表达的调控有关。事实真是这样。对真核生物尤其是这样,因为真核生物是多细胞的,它的机体是按照一定的时空程序生长、发育的,DNA 的许多序列都与此有关。按照一定时空程序的生长发育无疑属遗传信息,因此,DNA 所携带的遗传信息至少包括两方面:一是规定体现生物机体功能的各种蛋白质结构的信息(蛋白质基因),另一是规定基因选择性表达的信息。这两类信息的性质不同,方式也不同。

对蛋白质基因来说,DNA 的核苷酸序列通过mRNA 规定蛋白质的氨基酸序列,两者为线性关系,同时信息的转换要通过蛋白质基因以外的因子,如tRNA、核糖体等来完成。而有关基因表达调控的信息则与此不同,它是通过DNA 有关序列所形成的特异性结构与有关蛋白质因子的结合来体现的。

这一领域正在成为分子生物学发展的一个重要前沿。

第十三节 DNA 的限制图谱和序列分析

DNA 的核苷酸序列是其结构和功能的基础,因此,序列分析就成了分子生物学的一项重要的基础工作。不过,这一工作只是在发现限制性内切酶和DNA 重组技术及电泳技术得到长足进步后才获得现实意义。

一、限制性内切酶

限制性内切酶(restriction endonuclease)的发现是分子生物学发展中一个重大事件,它对核酸研究做出了重大贡献。

在很长时间内,人们认为所有的核酸内切酶相对而言都是非专一性的。然而不久前在原核细胞内发现了一类新的核酸内切酶,它们仅切断特定序列中的磷酸二酯键,这些酶称为限制性内切酶。

当一种外来DNA 进入细菌的细胞,它们就会被其中的核酸酶破坏,但细菌自身的核酸却不会受到这些酶的作用。这是细菌对入侵DNA 可能干扰其遗传信息的一种防卫。这一过程涉及到外来和自身DNA 的识别和外来DNA 的降解,这称做限制-修饰系统(restriction-modification system)。限制-修饰系统最早是由噬菌体的感染效应发现的:由一种细菌释放的噬菌体能够感染同种的细菌,但不能感染其它种的细菌,因为它们在进入后者细胞后即被破坏(限制)了。

限制性内切酶大致可分为两大类。这两类限制性内切酶的主要区别在于：类型 II 的限制-修饰系统由两种酶组成，限制酶仅使 DNA 降解，而 DNA 的修饰由另一种甲基化酶承担；类型 I 的这两种活力存在于同一酶分子上。不仅如此，两类酶的产物，辅助因子等也都不同(表 4-6)。

表 4-6　两类限制性内切酶

	类型 I	类型 II
裂　解	随机	位点专一性
产　物	异质的	专一片段
DNA 的识别	位点专一	位点专一
对 ATP 的需求	是	否
对 SAM 的需求	是	否
对 ATP 酶的依赖	是	否

类型 II 限制-修饰系统中的甲基化酶能使细菌自身 DNA 的专一性腺嘌呤或胞嘧啶残基甲基化。这些高度专一的甲基化酶与原核细胞中生成 m^5C 和 m^6A 的较为通用的甲基化酶不同，它们仅识别和使特定序列中的有关碱基甲基化。这样经过甲基化的 DNA 就能被细胞中的限制酶所识别而免遭降解。DNA 的甲基化和切割都发生在其分子的识别序列中，亦即限制酶

表 4-7　不同限制性内切酶的识别序列

酶	识别位点（对称轴左侧）	识别位点（对称轴右侧）	酶	识别位点（对称轴左侧）	识别位点（对称轴右侧）
EcoR I	5′-GAA CTT	TTC AAG	*Hind* II	5′-GTPy GApu	PuAC PyTG
Hind III	AAG TTC	CTT GAA	*Hpa* I	GTT CAA	AAC TTG
Hpa II	CC GG	G G C C	*Hae* III	GG CC	CC GG
Acy I	GPuC CPyG	GPyC CPuG	*Afl* III	ACPu TGPy	PyGT PuCA
Aha II	GPuC CPyG	GPyC CPuH	*Ava* I	CPyC GPuG	GPuG CPyC
Cfr I	PyGG PuCC	CCPu GGPy	*Gdi* II	PyGG PuCC	CCG GGC
Hae II	PuGC PyCG	GCPy CGPu	*HgiC* I	GGPy CCPu	PuCC PyGG
Hgi II	GPuG CPyC	CPyC GPuG	*NscP* I	PuCA PyGT	TGPy ACPu
Alu I	AG TC	CT GA	*Asu* II	TTC AAG	GAA CTT
Cla I	ATC TAG	GAT CTA	*BseP* I	GCG CGC	CGC GCG

箭头表示切割位点；Pu＝嘌呤，Py＝嘧啶。所有序列上方一股的 5′ 端在左侧。

和甲基化酶都结合于同一序列上。不同限制酶的识别序列不同。因此，一种特定的细菌仅能获得针对其自身限制酶的保护，而不能免受具不同识别序列的限制酶的切割。类型Ⅱ限制酶对核酸研究特别有用。

类型Ⅱ限制酶的识别序列常是一组含4～6核苷酸的回文序列（表4-7），少数可达8核苷酸。这种序列显示出二重对称性，如：

5′-AAGCTT
TTCGAA-5′

这种序列，如绕其中心转动180°，它的碱基顺序完全相同。由于这种对称性，两股链上的碱基都可被甲基化；同样限制酶亦引起DNA双股的断裂，并且几乎是同时发生的。识别序列并不总是连续的，它们可以为一个或更多的非专一性碱基隔开。细胞中，DNA靶位点可以是全甲基化、半甲基化和非甲基化的。对于这些靶位点，甲基化酶和限制酶的作用不同。故其命运也不同（图4-46）。甲基化酶和限制酶对全甲基化的识别序列都不作用；限制酶对半甲基化序列虽不作用，而甲基化酶则可使其进一步甲基化生成全甲基化序列，这样就保证了复制成的半甲基化序列不被降解。

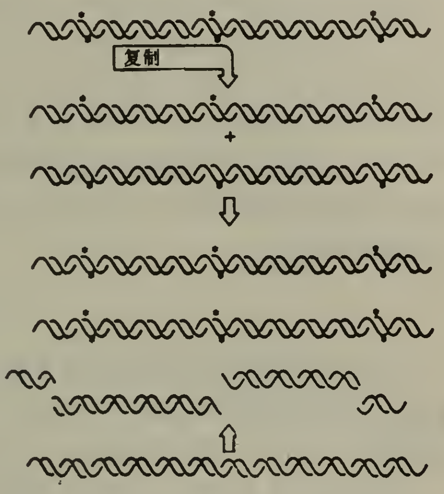


图4-46　DNA中甲基化和非甲基化序列
* 表示甲基化位点。

限制酶对DNA的切割点一般都位于识别序列内。多数情况下，DNA两股链的切割位点偏离识别序列的中心，故酶作用的结果会产生单股尾巴，这些单股尾巴如并列放置，相应的碱基间就通过氢键而缔合，故它们称为粘性末端（cohesive end）。有些酶沿同一位点切割双股链，结果就产生平整末端（blunt end）（图4-47）。也有一种情况是：两种不同的酶虽结合于相同的识别序列，但在不同的核苷酸间进行切割（即切点不同）。迄今发现的识别序列中，G和C常占优势。这一情况的存在是否由于鸟嘌呤和胞嘧啶残基间具有较强的配对能力则尚不清楚。近来，发现了一种其识别序列全由AT残基（TTTAAA）组成的酶（*Aha* Ⅲ）。

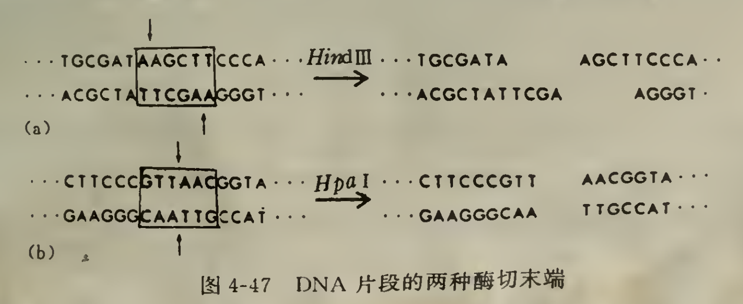


图4-47　DNA片段的两种酶切末端

由4种碱基所形成的某一组合要每几百碱基对才可能遇到一次，而6种碱基对的特定组合则需几千碱基对才能出现一次。因此，限制性内切酶切割所产生的DNA片段都具特定的长度。例如，*Eco*RⅠ的含6 bp的识别序列在SV40 DNA（5 243 bp）中仅出现一次，而*Hin*dⅢ则可在5处切割SV40 DNA。此外，存在于一种DNA中的序列在另一种DNA中常常缺如。例如，T7噬菌体DNA（39 936 bp）甚至完全不含*Eco*RⅠ的识别序列。

由限制性内切酶产生的是 DNA 的特异性片段。这些片段对建立 DNA 的限制图谱和重组体等都十分有用。

经限制性内切酶作用所产生的酶切片段可以方便地通过琼脂糖凝胶电泳分离。在琼脂糖凝胶中,DNA 片段的迁移速率依赖于它们的长度,迁移的距离随片段长度的增加而减少。电泳后,DNA 的位置可藉溴乙锭处理而显现为可见的条带。DNA 片段的大小则可通过已知分子量 DNA 的对照而确定(图 4-48)。分离后的 DNA 片段可以活性形态回收。

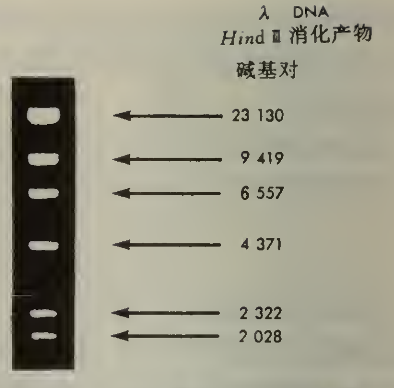


图 4-48 DNA 酶切片段的琼脂糖凝胶电泳分离

二、DNA 的限制图谱

DNA 限制酶切片段沿染色体或染色体片段的依次排列称做限制图谱(restriction map)。酶切片段可用选定的限制酶处理待测 DNA 获得,然后通过琼脂糖凝胶电泳分离,并根据泳动距离推算各自的分子量(泳动距离与分子量的对数成反比)。第一张这样的限制图谱就是 *Hind* Ⅱ,*Hind* Ⅲ 以及 *Eco*R Ⅰ 这样的限制酶处理 SV40 DNA 获得的(图 4-49)。*Eco*R Ⅰ 在 SV40 DNA 上只有一个切点,故它使 DNA 由环型转变为线型,图中标出的零点即是酶的切点。所得的线型病毒 DNA 再依次由 *Hind* Ⅱ,*Hind* Ⅲ 处理而降解成 11 个片段。DNA 片段的依次排列是通过 DNA 的不完全彻底酶解所得各个片段来实现的,不完全酶解所得的片段数少于完全酶解的,而多数片段的分子量则大于后者,由于酶在 DNA 上的切点是固定的,因此,酶切片段间必然会出现固定的重叠区。根据这些重叠区,完全酶解的片段就可沿染色体片段依次排列。由不同限制酶得到的图谱是不同的,最初使用的常是那些对待测 DNA 的切点较稀少的限制酶,因为这样可产生较少的便于分离的片段。

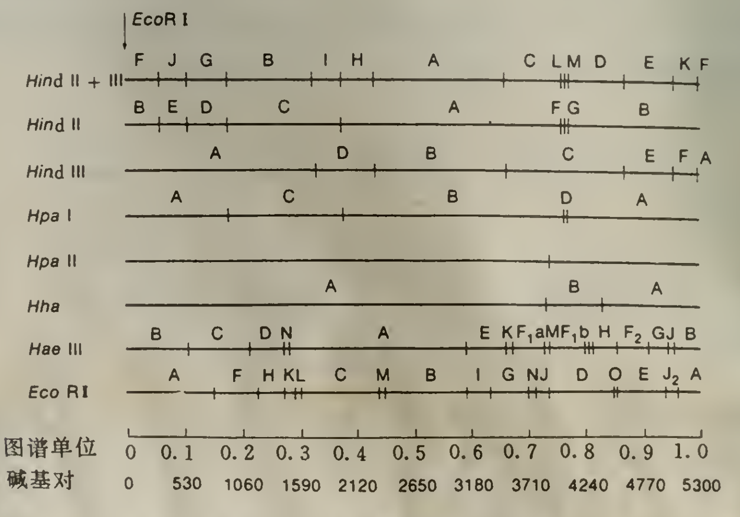


图 4-49 SV40 基因组的限制图

*Eco*R Ⅰ 对 SV40 DNA 只有一个切点,它使 SV40 DNA 转变为一线型分子。为了便于与其它限制酶切片段做比较,故把酶切片段从 *Eco*R Ⅰ 的切点(作为 0 点)开始展现为开环型。每一直线各代表不同酶的切点和形成的片段。

酶的识别位点在DNA中的多寡决定所产生的酶切片段的数量。对长DNA言，人们希望找到在该DNA中识别序列较少的酶，以免酶切片段过多。现在已可通过巧妙利用两种高度专一性的甲基化酶为限制酶 *Dpn* I 构建包括8和10 bp的人为延长的识别序列。这种人为延长的识别序列大大方便了工作的开展，因为即便是8 bp的识别序列在DNA中至少也要每65 000(4^8)bp才有机会遇到一次，故延长的识别序列就可能使一个很大的DNA分子仅断裂成较少的片段。构建这种人为延长的识别序列之所以成为可能是由于利用了限制酶 *Dpn* I 的一种独特性质，即它仅切割含甲基化腺嘌呤的限制位点。

$$\begin{array}{l} G^{m}\overset{\frown}{A}\ TC \\ C\quad T^{m}AG \end{array}$$

因此，*Dpn* I 一般不能切割大多数生物的DNA(例如，脊椎动物DNA)。但这种DNA如以一种其识别序列与 *Dpn* I 的识别序列相重合的甲基化酶处理，DNA分子中适当的腺嘌呤残基就被甲基化，这样的序列也就随即可为 *Dpn* I 所作用。例如，*M.Taq* I 能使识别序列中顺向重复的TCGA发生甲基化，从而产生一个含8 bp的 *Dpn* I 识别序列(图4-50)。*M.Cla* I 可以识别两个部分重叠的ATCGA序列，并使其甲基化，故从而构建出含10 bp的 *Dpn* I 识别序列(图4-50)。

5′-TCGATCGA
AGCTAGCT-5′
↓甲基化酶
5′-TCG^{m}A TCG^{m}A
AGC T^{m}AGC T-5′
↓限制性内切酶
5′-TCG^{m}A-3′ 5′-TCG^{m}A
^{m}AGC T-5′ 3′-AGC T-5′

(a)

5′-ATCGATCGAT-3′
3′-TAGCTAGCTA-5′
↓甲基化酶
5′-A TCG^{m}A TCG^{m}AT-3′
3′-T^{m}AGC T^{m}AGC TA-5′
↓限制性内切酶
5′-A TCG^{m}A-3′ 5′-TCG^{m}AT-3′
T^{m}AGC T-5′ ^{m}AGC TA-5′

(b)

图4-50 人为延长的限制酶识别序列
(a)含8 bp的延长识别序列；
(b)含10 bp的延长识别序列。

这种实际上很稀少的识别序列对构建含多于10^5 bp的DNA的限制图谱是十分有用的。虽然直至不久前，人们还无法分离这些构建的识别位点所产生的很大DNA片段。但现在，由于采用琼脂糖凝胶脉冲电泳，高至10^6 bp的DNA也可进行常规处理了。现在大肠杆菌染色体(4×10^6 bp)已经被裂解为20个片段，它的限制图谱可望即将完成。

三、DNA的序列分析

随着限制性内切酶的发现和DNA重组技术的建立，紧接着就出现了两种确定DNA片段中4种碱基顺序的有效方法，即Maxam和Gilbert法及Sanger法。Maxam和Gilbert法亦称化学法，Sanger法亦称酶法，以后又在这两个方法的基础上提出了不少改进法。

Maxam和Gilbert法的基本原理是使待测DNA的单股链，通过选定条件下专一性试剂(如硫酸二甲酯)的作用，裂解成较小的片段，由于专一性试剂只作用于特定碱基，因此，裂解产生的是具一定长度的片段。图4-51中的DNA单股链，如使试剂仅专一性地破坏C，不完全的裂解反应可产生图中所表示的几种产物。由于单股链仅在5′端标记，故放射自显影能检测的仅是保留原来5′端的片段。如裂解反应进行到底，具放射性的裂解产物就只有一种，结果就不能检测出其它胞嘧啶残基所在的位置。故在测序中，使反应不进行到底是重要的。Maxam和Gilbert测序法的要点是：

(1)取待测 DNA 的一股链，使其一端(一般为 5′端)用^{32}P 标记。

(2)样品分为 4 份，分别在选定条件下与专一性试剂作用。每一份样品各生成一套长短不等的放射性片段。

(3)应用凝胶电泳分离每一反应的产生，放射自显影后，根据按长短排列、具特定末端碱基的 DNA 片段就可读出待测 DNA 的碱基序列(图 4-51)。现在应用凝胶电泳可以分离仅相差一个核苷酸的 DNA 片段，在同一凝胶板上可解析出 300 以上的条带。这一方法最初被用于测定细菌 DNA 中 lac 操纵子调节区的序列。其后不久就实现了以一个科学家的力量在一年时间内就完成了质粒 pBR322(4 362 bp)的序列分析。

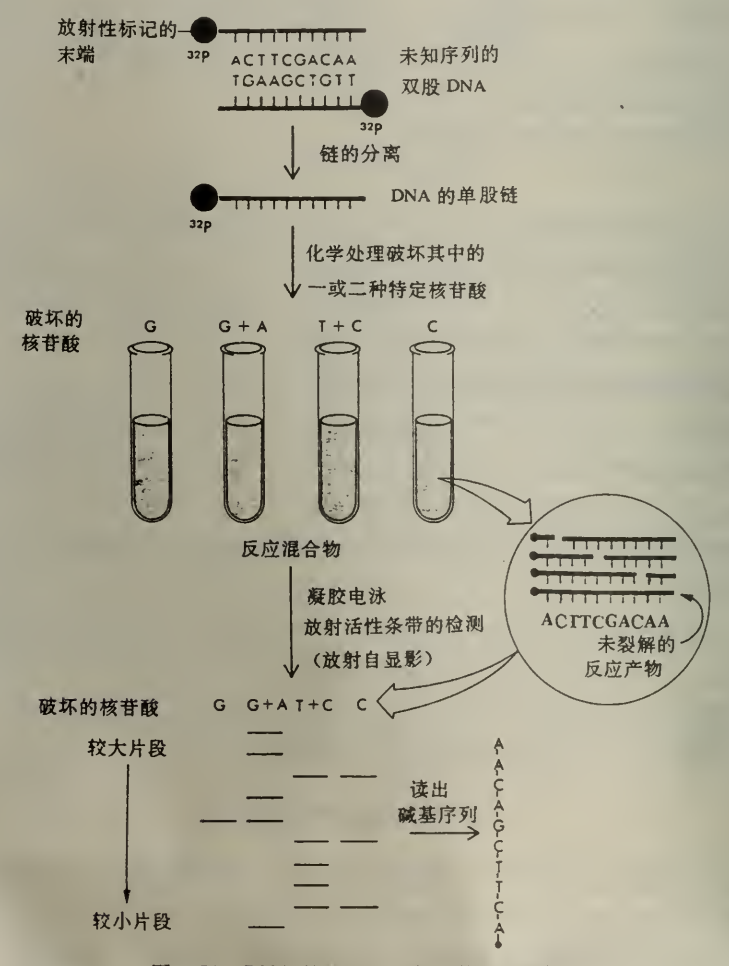


图 4-51　DNA 的 Maxam 和 Gilbert 测序法

第二种广泛使用的 DNA 测序法是 Sanger 法。这两种方法的基本区别在于：第一种方法是利用专一性试剂使待测 DNA 在特定的碱基处裂解为长短不等的 DNA 片段；第二种方法是在特定 2′,3′-二脱氧核苷酸(2′,3′-dideoxy nucleotide)的存在下用合成方法生成长短不一、具特定末端的 DNA 片段。由于二脱氧核苷酸不再能延伸生成 3′,5′-磷酸二酯键，合成反应也就在该处停止，故这一方法也称为双脱氧法。Sanger 法的要点(图 4-52)是：

(1)选取待测 DNA 的一股链为模板，使 5′端标记的短链引物互补于模板的 3′端。

(2)以样品分成 4 份，分别以 4 种脱氧核苷酸和相应于其中一种的二脱氧核苷酸为底物，加入 DNA 聚合酶 I 引发 DNA 合成。二脱氧核苷酸与正常核苷酸发生竞争，故合成反应在二脱氧核苷酸掺入处停止，结果生成一套长短不等的片段。

(3)应用凝胶电泳分离每一合成反应的产物。根据所得条带的位置，读出待测 DNA 的碱基序列。

这一方法一经问世就受到人们的关注，很快就测出了 T4 噬菌体 DNA(5 577 bp)的完整序列，其后又建立了噬菌体 λDNA(48 513 bp)的完整序列。这一方法正得到越来越多的应用。

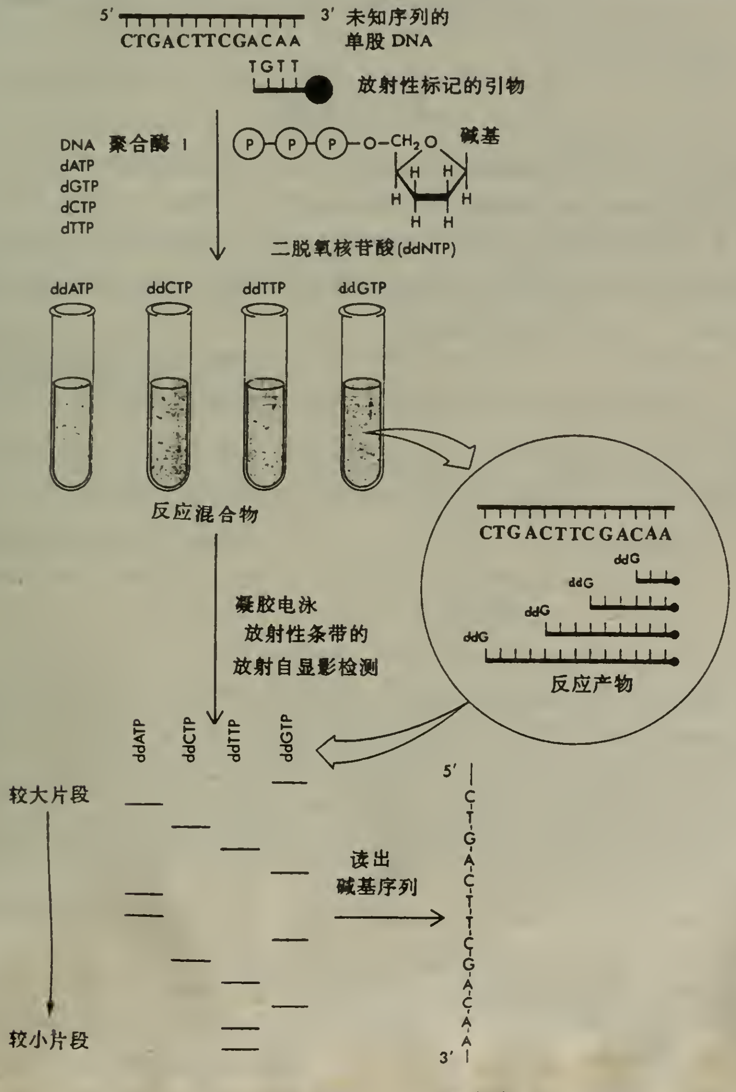


图 4-52　DNA 的 Sanger 测序法

DNA 的测序法现在已日益走向成熟。在上述两种方法的基础上已发展出不少改进的方法。同时，应用计算机来贮存 DNA 的碱基序列亦成为一种常规方法。当碱基序列由凝胶读出时即可直接输入计算机中，这就避免了手工抄录很长序列时常易发生的差错。一旦输入计算机，限制酶的识别位点，RNA 合成的起始和终止信号，反向回文结构，可能形成 Z-DNA 的优

势片段(交替嘌呤-嘧啶区)和其它已知DNA 的同源序列等就都可由扫描获得。DNA 的碱基序列也可转读成蛋白质的氨基酸序列。实现这些工作的高效软件程序已可由多种来源获得,不过现有的巨大数量的DNA 序列方面的信息则需要采用更为强大的计算机。

DNA 序列测定本身可以藉采用强大的计算机而加速,但建立长序列的限制图谱所需的时间则远比测序为多。因此,人们希望测定许多可能重叠的DNA 片段,然后利用计算机寻找能据以使DNA 片段沿各自染色体依次排列的重叠序列。不过这一方法只是在二脱氧测序法的M13 改进法(参阅第十三章)出现后才变得具有实际意义。M13 法使序列测定变得十分简便。可使工作以在几年前还难以想象的速度(每周约2 000 bp)进行。不久,一些很大染色体片段的碱基序列都可能最终被测定。

参考文献

1. 白春礼,叶坚,龚立三. DNA 变异结构的扫描隧道显微镜研究. 科学通报. 1990. 24:1 841～1 842.
2. Blackburn, E. H. Structure and Function of Telomeres. Nature. 1990. 350:569～573.
3. Dickerson, R. E. The DNA Helix and How It Is Read. Sci. Am. 1983. 249:86～102.
4. Frank-Kamenetskii, M. D. and Mirkin, S. M. Triples DNA Structure. Annu. Rev. Biochem. 1995. 64: 65～95.
5. Kornberg, A. and Baker, J. A. DNA Replication. W. H. Freeman and Company, New York:1992.
6. Lehninger, A. L., Nelson, D. L. and Cox, M. M. Principles of Biochemistry, Worth Publishers, Inc. New York:1993.
7. Mainwaring, W. I. P., Parish, J. H. and Pickering, J. D., Mann, N. H. Nucleic Acid Biochemistry and Moiecular Biology. Blackwell Scientific Publications, Oxford London:1982.
8. Mirkin, S. M., Lyamichev, V. I., Drushlyak, K. N., Dobrynin, D. N., Filippov, S. A. and Frank, Kamenetskii, M. D. DNA H Form Requires a Homopurine-homopyrimidine Mirror Repeat Nature. 1987. 33: 495～497.
9. Sanger, W. Principles of Nucleic Acid Structures. Springer Verlag. New York:1984.
10. Schultze, P., Smith, F. W. and Feigon, J. Refined Solution Structure of the Dimeric Guadruples Formed from the Oxytricha Telomeric Oligonucleotide d(GGGGTTTTGGGG). Structure 1994. 2: 221～233.
11. Voet, D. and Voet, J. D. Biochemistry John Wiley and Sons. New York:1990.
12. Wells, R. D. Unusual DNA Structures. J. Biol. Chem. 263:1 095～1 098.
13. Wang, Y. and Patel, D. J. Solution Structure of the Human Telomeric Repeat d[$AG_3(T_2AG_3)_3$]G-tetraplex Structure. 1993. 1:263～282.
14. Watson, J. D. and F. H. C., Crick. Molecular Structure of Nucleic Acids: A Structure for Deoxynucleic Acids. Nature. 1953. 171:737～740.
15. Watson, J. D. and F. H. C., Crick. Genetical Implication of the Structure of Deoxyribonucleic Acid. Nature. 1953. 171: 964～968.
16. Watson, J. D., Hoprins, N. H., Roberts, J. W., Steitz, J. A. and Weiner, A. M. Mole cular Biology of the Gene Benjamin/Cummings Publishing Company, Menlo Park:California. 1987.

(张玉麟)

第五章　DNA 的复制和修复

DNA 是遗传信息的载体，细胞传代时，DNA 必须忠实地复制才能使子细胞含有相同的遗传信息，从而保持物种的稳定。这就是为何 DNA 的双螺旋结构提出时，遗传学家所以如此感兴趣的原因，因为 DNA 双股链间的互补关系为复制提供了使人信服的基础。

DNA 通过其双股链间的互补关系所实现的复制，初看起来似是一个直截了当的过程，但如再加审视，DNA 复制的复杂性也是显而易见的。例如，DNA 的复制是在单链上进行的，而 DNA 在细胞中是以高级结构存在的，那么 DNA 的结构在复制前后究竟发生怎样的变化呢？在细胞周期中，DNA 复制只发生在 S 期，那么 DNA 复制的起始与终止是怎样调控的呢？DNA 是一些巨大的分子，那么它的复制是从分子的固定位点开始还是随机的呢？为了物种的稳定，DNA 的复制必须准确无误，那么复制所要求的这种高度保真性是怎样达到的呢？如此等等，不一而足。实际也证明，DNA 复制是一个涉及多方面问题的复杂过程，是一个由多种酶催化和有多种蛋白质参与的受到精密调控的过程。

DNA 复制要求能保持遗传信息的完整，但复制的精确性毕竟不是绝对的。同时，DNA 还会受到各种因素的损伤，这些作用虽然十分缓慢，但由于生物对此的耐受力十分低，故仍可引起严重后果。为了保持物种的延续，细胞通过多种修复系统使损伤减少到最低限度。DNA 是细胞中唯一具修复系统的生物大分子。

第一节　DNA 复制概貌

无论是染色体的数目或是染色体 DNA 的大小，原核细胞都要比真核细胞为小。同时，原核细胞生活周期短，容易在人工条件下培养，原核细胞容易获得各种突变体，而原核细胞中的大肠杆菌又是在各方面研究得最为透彻的一种生物。这些就使原核细胞成为研究 DNA 复制（DNA replication）的一种良好的实验系统。此外，基因组比原核细胞更小的噬菌体对前者的感染又为这方面的工作提供了很好的研究途径。因此，我们目前所掌握的有关 DNA 复制的知识主要来自原核细胞 DNA。

一、DNA 复制的半保留性

DNA 如何以其所携带的遗传信息通过复制传递给后代可能有三种方式：全保留（conservative）、半保留（semiconservative）和分散（dispersive）复制（图 5-1）。

实验证明，DNA 是按半保留方式复制的，这一著名实验是 Meselson 和 Stahl 在 1958 年完成的。他们先以大肠杆菌在含 ^{15}N（NH_4Cl 形态）的培养基中繁殖多代，使嘧啶和嘌呤碱基中的 ^{14}N 全被置换为 ^{15}N。收集大肠杆菌，分离其中的 DNA，然后再进行 CsCl 平衡密度梯度离心，这时 DNA 形成一单独的，浮力密度为 1.724 g/ml 的条带。和对照（浮力密度为 1.710 g/ml）相比，DNA 浮力密度（buoyant density）的这种增加是由于 ^{15}N 置换了 ^{14}N。现在再以在 ^{15}N 氮源

中培养的大肠杆菌转移到含^{14}N的培养基中传代，每隔一定时间取样，分离其DNA并进行密度梯度离心。这时，由浮力密度不同所产生的DNA条带显示了规律性变化（图5-2）。这种规律性变化只能由DNA的半保留复制得到解释。在零代细胞中，DNA双股链中氮的分布为$^{15}N/^{15}N$，其浮力密度为1.724 g/ml；在第一代细胞中，DNA双股链中氮的分布由$^{15}N/^{15}N$转变为$^{15}N/^{14}N$，其浮力密度也由1.724转变为1.717 g/ml；在第一代以后的细胞中，DNA双股链中氮的分布有两种，即$^{15}N/^{14}N$和$^{14}N/^{14}N$，条带扫描曲线的峰也分裂为二，分别相应于浮力密度1.717和1.710 g/ml。随着细胞传代的进行，双链均含^{14}N的DNA的比重也越来越高。这一实验证明，复制后的DNA是由一条亲代链和一条子代链组成的，即复制是按半保留方式进行的。

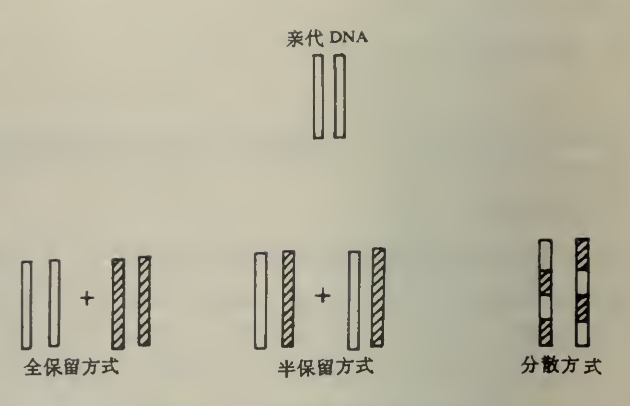


图5-1　DNA复制的可能方式

□代表原来的DNA；▨代表新合成的DNA。

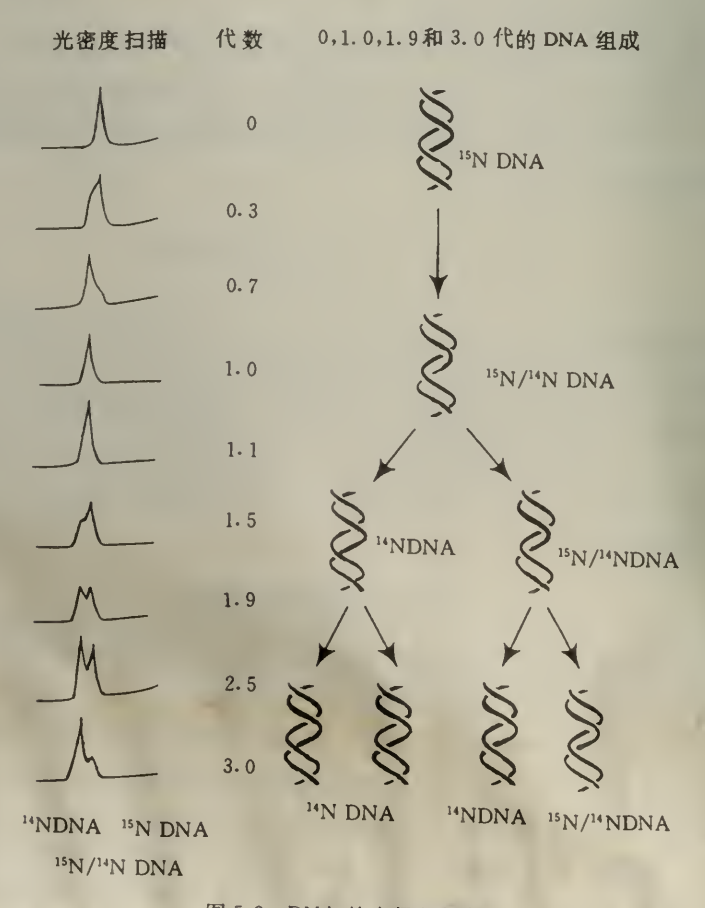


图5-2　DNA的半保留复制

在这一实验中所使用的技术称为密度标记DNA(density labeled DNA),它在DNA的复制、修复和重组研究中都有广泛的应用,当然标记的方法也可不同,如以5-溴脱氧尿苷来代替胸腺嘧啶核苷就可获得DNA的更大密度增量。

二、复制过程的顺序性

复制的半保留性仅说明,子代DNA分子中的一股链来自亲代;另一股链是新合成的,但它并未指出复制的进行过程,即双螺旋DNA是先全部解开成两股单链,然后分别作为模板合成新链,还是边解开边合成的。

许多实验证明,DNA的复制是按一定的程序进行的。SV40 DNA复制早期的电镜照片指出,分子中除了已复制的部分外,还存在着呈超螺旋的未复制部分(图5-3)。复制从分子的称

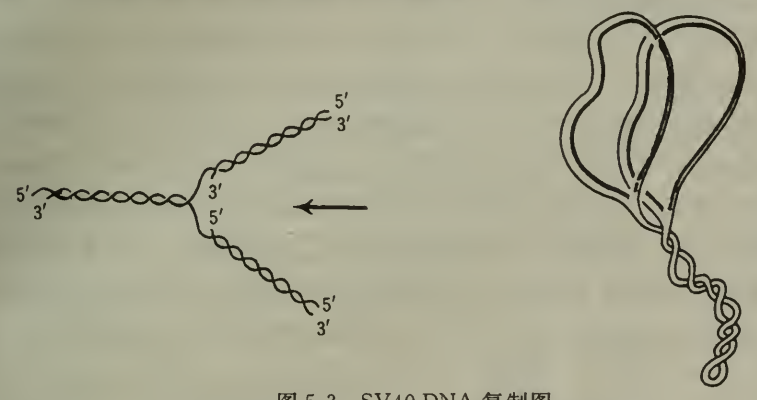


图5-3 SV40 DNA复制图

右图:处在复制的SV40 DNA分子;左图:复制叉。

为原点(origin)的特定位点开始,然后向未复制部分逐步扩大。DNA分子中正在进行复制的分叉部位称为复制叉(replication fork),它由两股亲代链及在其上新合成的子链构成。DNA的正在复制的部分在电镜下观察起来犹如一只眼睛,称为复制眼(replication eye)。对环形DNA言,复制眼的存在就使其成为图5-4所示的θ结构(θ structure)。复制眼本身并不能说明它是由一个或是二个复制叉形成的,亦即复制是单向还是双向的(图5-5)。事实说明,单向或是双向复制都是存在的,但多数DNA的复制是双向的,SV40DNA的复制就是一例。SV40的基

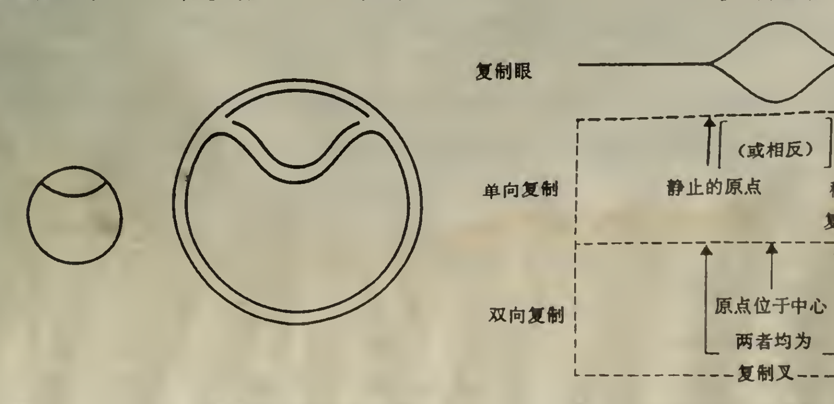


图5-4 环形DNA复制所形成的θ结构

图5-5 DNA可能的单向或双向复制

因组是一个含 5 224 bp 的环形 DNA，它仅含限制性内切酶 *Eco*R Ⅰ 的一个切点，

$$\begin{array}{c} \downarrow \\ 5'-GAATTC-3' \\ 3'-CTTAAG-5' \\ \uparrow \end{array}$$

故当复制中 SV40 DNA 群体用 *Eco*R Ⅰ 处理时，它们就转变为具不同大小复制眼的线型分子。在这样所得的不同线型分子中，复制眼两侧的 DNA 链的长度随复制眼的扩大而相应缩短，这一情况有力地说明 SV 40 DNA 的复制是由原点开始然后双向进行的。

三、DNA 的半不连续复制

双螺旋 DNA 的两股链的极性是相反的，因此，在复制叉两股亲代链上合成的子链的极性也应相反。据此就产生了一个问题，即一股子链的 DNA 合成如沿 5′→3′ 方向延伸，则另一股链上 DNA 的合成似就应按 3′→5′ 方向进行。事实是，至今没有能找到 DNA 合成能按 3′→5′ 方向进行的证据。但如双股链上 DNA 的合成都按 5′→3′ 方向进行，由于双螺旋 DNA 是逐步解旋的，则一股链上的 DNA 合成可以是连续的；另一股链上的合成只能是不连续的。Okazaki（冈崎）证明情况确是如此，它用〔^{3}H〕胸腺嘧啶脱氧核苷脉冲标记大肠杆菌培养物，这时优先标记的是新合成的 DNA。沉降分析证明新合成的大多是一些约含 1 000 核苷酸的短片段。如以培养物再在非放射性培养基中保温，标记就进入高分子量 DNA 中，这些实验就证明 DNA 的复制是半不连续的（semidiscontinuous），即一股模板上的 DNA 合成是连续的；另一模板上的合成是不连续的。同时，复制叉一股链上的 DNA 合成总是比另一股超前一步，称为先导链（leading strand），另一股的复制要滞后一步，称为后随链（lagging strand）。在不连续合成中形成的短 DNA 片段称为冈崎片段（Okazaki fragment），它们随后连接成大片段（图 5-6）。

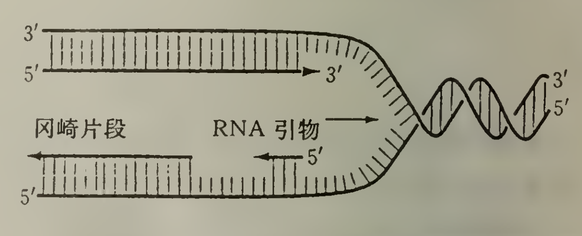


图 5-6　DNA 的半不连续复制

四、复 制 单 位

各种生物基因组的大小不同，尤其是原核和真核生物差异更大。大肠杆菌的基因组是单个环型双股 DNA，含 3～9×10^{6} bp，它仅含一个原点，复制一次约需 40 min。真核生物基因组要比原核的大许多，它复制时每秒所合成的核苷酸数比原核少（表5-1）。如真核生物的每一

表 5-1　不同细胞的复制参数

	大肠杆菌	人*
DNA 含量，每一细胞的核苷酸数	3.9×10^{6}	约 10^{9}
复制叉的移动速率（μm/min）	30	3
DNA 复制速率，核苷酸数/s/复制叉	850	60～90
每一细胞的复制原点数	1	10^{3}～10^{4}
整个基因组复制所需的时间（h）	0.67	8
细胞分裂所需的时间（h）	0.33	24

* 人 DNA 的复制参数来自 Hela 细胞。

DNA 分子也仅含一个原点，细胞所有 DNA 复制一次就需几个星期。实际情况是，真核生物 DNA 含许多复制原点，使整个 DNA 分子可分为多个可同时复制的单位，从而使 DNA 复制可以在细胞周期的 S 期中完成。DNA 的这种能独立复制的单位称为复制子(replicon)。

复制子概念可以通过大肠杆菌的染色体和附加体的关系来说明。大肠杆菌含有一些称为质粒(plasmid)的染色体外遗传因子(extrachromosomal genetic element)。质粒是一些小的环形 DNA，在大肠杆菌细胞中，染色体 DNA 和质粒 DNA 分别独自复制。质粒中，有些称为附加体(episome)的 DNA 能整合(integration)进染色体 DNA，这时它的复制就和后者同步进行。大肠杆菌有一称为温度敏感型(temperature sensitive type，ts)的突变体，它改变了 DNA 复制的引发。这种突变体在 30℃能正常复制，但在 42℃却不能引发 DNA 复制。在未整合前，染色体 DNA 的复制是双向的，附加体的复制是单向的；整合后，附加体的复制就受染色体 DNA 复制的控制。当温度提高至 42℃时，由染色体 DNA 原点开始的双向复制停止，代之的是由附加体原点开始的单向复制(图 5-7)。从这一现象可以看出，DNA 复制的起始是由一组基因控制的，这些基因包括复制原点和其它的结构基因，复制即是在结构基因的表达产物作用于原点而引发的。上述的这些基因以及在其控制下的 DNA 区段统称为复制子。原核 DNA 构成一个复制子，而真核 DNA 则构成多个复制子。

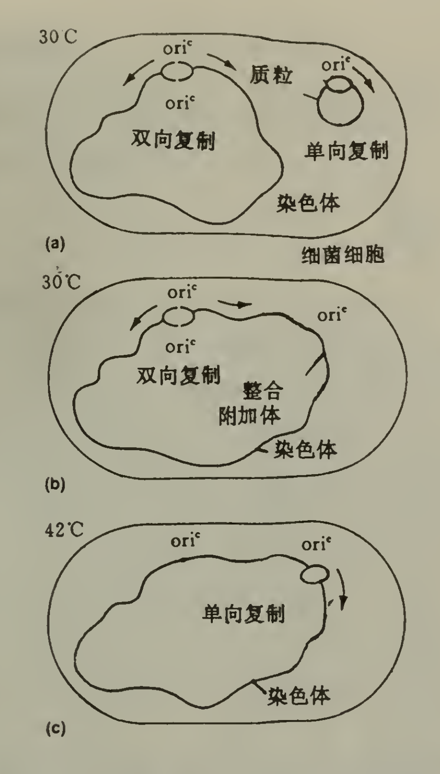


图 5-7　附加体和染色体 DNA 复制的关系

从表 5-1 还可以观察到一个表面上异常的现象，即细菌的细胞数每 20 min 增加一倍，但其染色体DNA 的一轮复制却需 40 min。实际情况是，在快速生长的细胞中，DNA 在完成一轮复制前即开始了新一轮的复制。这些实验有力地表明，DNA 复制的控制主要是在起始水平上实现的。

第二节　DNA 的复制酶和相关蛋白

DNA 复制是一个涉及许多酶和蛋白的系统。尤其是真核生物 DNA 复制更为复杂，难以用较短的篇幅阐述清楚。鉴于原核生物 DNA 复制的相关酶和蛋白研究得较为深入，以及原核和真核生物 DNA 复制的一些基本原理相同，所以，本章就着重讨论原核生物 DNA 复制，而真核生物方面的情况就留待在适当的章节逐步深入。

DNA 复制是一个复杂过程，它包括双螺旋和超螺旋的解旋和重新形成，复制的起始和调控，模板上新 DNA 链的合成，复制的终止等等。复制是在许多相关酶和蛋白的协同作用下运行的。如按照在复制过程中出现的先后，它们包括：①DNA 旋转酶；②使 DNA 双股链在复制叉分离的蛋白质；③在复制前防止 DNA 链退火的蛋白质；④合成 RNA 引物的酶；⑤DNA 聚合酶；⑥除去 RNA 引物的酶；⑦使冈崎片段共价连接的酶等等。

一、核苷酸的聚合和 DNA 聚合酶

1957 年，Kornberg 报道在大肠杆菌抽提液中发现了能催化 DNA 合成的酶。

$$(dNMP)_n + dNTP \rightleftharpoons (dNMP)_{n+1} + PP_i$$

DNA 片段　　　　延伸的 DNA 片段

上式中的 dNMP 代表脱氧核苷-磷酸，dNTP 代表脱氧核苷三磷酸。这一反应需要加入短的 DNA 片段，否则不能进行；如加入的为完整的线型 DNA 或是环型 DNA，反应同样不会进行。实际情况是，反应是在加入的 DNA 一股链的 3′-末端延伸；另一股链决定掺入核苷酸的顺序(图 5-8)。加入的 DNA 片段既提供了反应的引物(primer)，又提供了合成反应的模板(template)。核苷酸的聚合是一亲核反应，引物 3′-末端的羟基向进入的 dNTP 的 α 磷酸基进行亲核进攻，而形成磷酸二酯键。这是一个可逆反应，它还需焦磷酸酶(pyrophosphatase)的作用使焦磷酸转变为正磷酸从而驱使反应趋向完成。这表明 DNA 中每掺入一个核苷酸需消耗两个高能磷酸键。这一催化核苷酸聚合的酶根据发现者称为 Kornberg 酶。以后，另外两种原核生物 DNA 聚合酶相继发现，故 Kornberg 酶就称为 DNA 聚合酶 Ⅰ (Pol Ⅰ)，而另两种分别称为 DNA 聚合酶 Ⅱ 和 Ⅲ (Pol Ⅱ 和 Pol Ⅲ)。三种 DNA 聚合酶中，聚合酶 Ⅲ 虽然是真正的复制酶(replicase)，但聚合酶 Ⅰ 发现的历史最长和在应用上的重要性，故是研究得最为深入的一种。同时，对它的讨论也有助于对其它聚合酶的认识，因此，对 DNA 聚合酶的讨论就从 DNA 聚合酶 Ⅰ 开始。

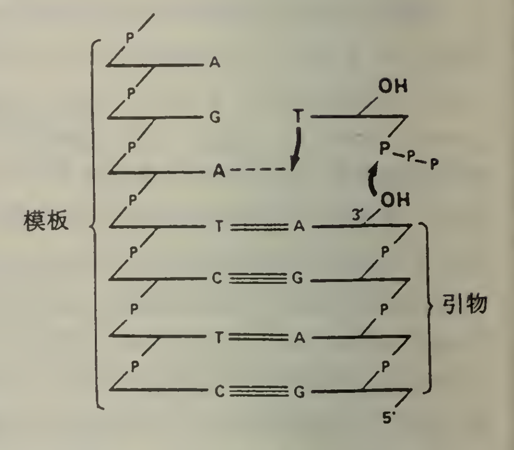


图 5-8　核苷酸的聚合反应

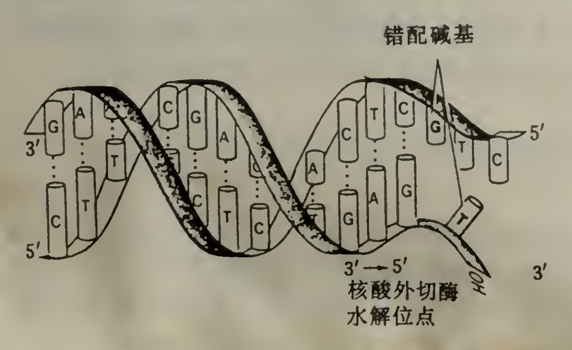


图 5-9　Pol Ⅰ 的 3′→5′ 核酸外切酶活力

DNA 聚合酶 Ⅰ 是一单体蛋白，分子量为 109 kD，共含 928 个氨基酸残基。DNA 聚合酶 Ⅰ 具有一些与一般酶不同的特性。DNA 聚合酶 Ⅰ 催化脱氧核苷酸的聚合，但脱氧核苷酸是按照一定顺序掺入的。这种专一性来自进入的脱氧核苷酸与模板间的碱基配对，而不是酶对碱基的直接识别。当某一脱氧核苷酸掺入 DNA 后，酶也并不立即与模板脱离，而是继续作用，这称为持续性(processitivity)。Pol Ⅰ 每一次与模板结合可使 10 或更多的脱氧核苷酸掺入 DNA，这加快了反应的速率。Pol Ⅰ 除了聚合酶活力外，还具有 3′→5′ 核酸外切酶活力和 5′→3′ 核酸外切酶活力。Pol Ⅰ 3′→5′ 核酸外切酶活力，是由具游离—OH 的未配对 3′-末端核苷酸所激活。若一个不能形成碱基配对的错配脱氧核苷酸掺入延长中的DNA链末端，聚合酶活力就被抑制，而3′→5′ 核酸外切酶活力被激活，结果错

配对的核苷酸就被切除(图 5-9)。聚合酶活力随即使复制重新开始。所以,Pol I 对 DNA 合成具校对作用。Pol I 的 5′→3′核酸外切酶活力则作用于具切口的双螺旋 DNA,并在切口以外的配对区切割,使具 5′-端的 DNA 链降解为单核苷酸或寡聚核苷酸(图 5-10)。Pol I 的聚合酶和 3′→5′及 5′→3′核酸外切酶三种活力存在于同一多肽链中(图 5-11)。当以枯草杆菌蛋白酶或胰蛋白酶处理时,Pol I 就裂解为大小不同的两个活性片段。较大的 C 端片段(残基 324～928)具有聚合酶和 3′→5′核酸外切酶活力,也称为 Klenow 片段;较小的 N 端片段(残基 1～323)具有 5′→3′核酸外切酶活力(图 5-11)。结构分析指出,Klenow 片段又含一大(残基 521～928)一小(残基 324～517)两个结构域。Klenow 片段较大结构域(图 5-12)的螺旋 I,O 间为一裂隙,DNA 即结合于此,螺旋 I 和 H 间的这部分结构则形成一可活动的盖,使所结合的 DNA 不易脱开。这一结构特点可用以解释 Pol I 作用的持续性。

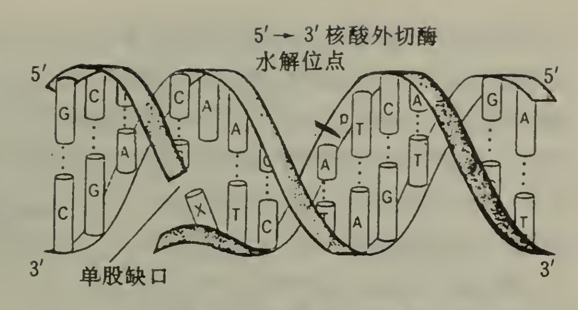


图 5-10　Pol I 的 5′→3′核酸外切酶活力

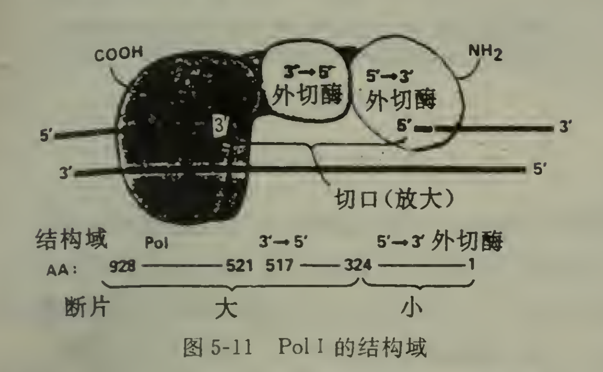
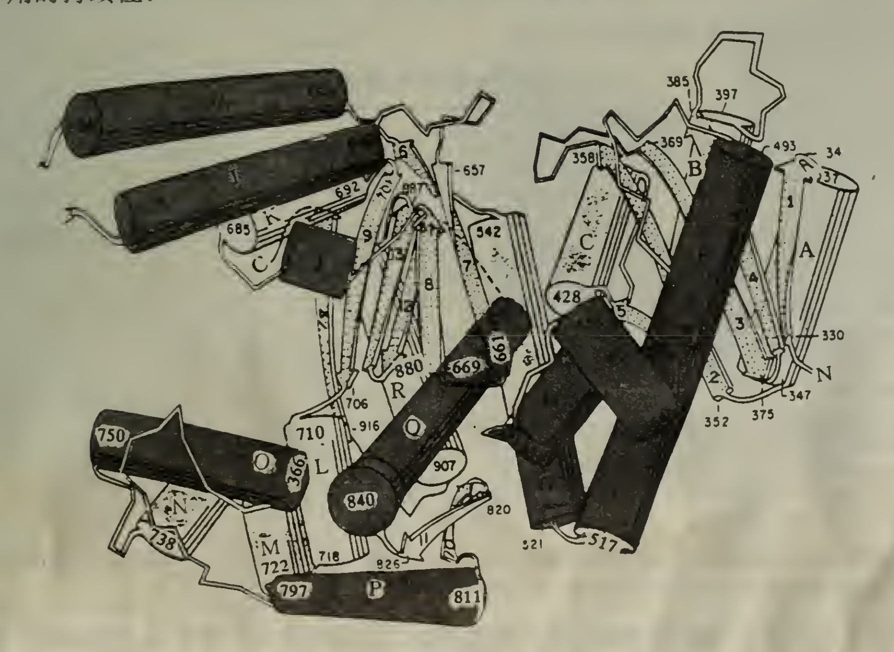


图 5-11　Pol I 的结构域


图 5-12　Pol I　Klenow 片段的结构

Pol I 是 1957 年发现的,人们曾一度认为它即是寻找的复制酶。可是,随后的许多事实表明情况并不如此。例如,它每分钟只能使 600 核苷酸掺入 DNA,实际测定数字要比这约大 20

倍。遗传分析也表明，大肠杆菌的复制还与其它基因产物有关。以后，Cairns 等在1969年发现，大肠杆菌一种突变体的抽提液，虽然仅显示出<1%的 Pol Ⅰ 活力，但该突变体却能以正常的速率繁殖。与此同时，该突变体对紫外辐射和化学诱变却表现出高度敏感。这表明复制可能主要由其它的聚合酶负责，而 Pol Ⅰ 的主要功能可能与DNA的修复有关。事实证明正是如此。

DNA 聚合酶Ⅰ的生理功能主要是DNA的修复。DNA在机体中会受到紫外辐射、化学诱变等许多物理和化学因素的作用而发生改变或损伤。如不加修复，机体就难于适应，或产生遗传上的突变，甚或死亡。细胞中存在多种DNA修复系统，它们的核酸内切酶活力会在DNA受损伤部位的5′端进行切割产生单链切口，后者就激活 Pol Ⅰ 的5′→3′核酸外切酶活力，从而切除DNA的受损伤部分，Pol Ⅰ 的聚合酶活力随即开始工作，使DNA链向3′端延伸，这称为切口位移(nick translation)(图5-13)。在体外，人们可利用切口位移制备高放射活性的DNA序列。DNA上的切口可用少量胰核糖核酸酶处理来产生，在放射性标记dNTP的存在下，切口位移即生成原来切口3′端的DNA序列。

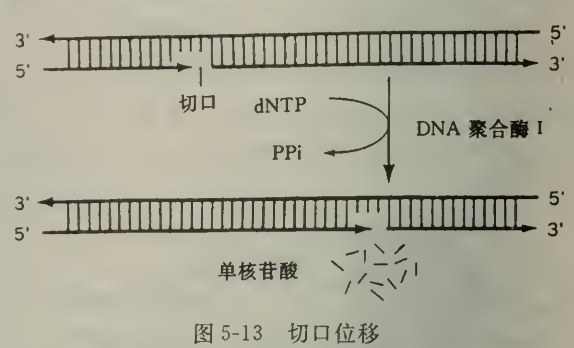


图5-13　切口位移

Pol Ⅰ 的5′→3′核酸外切酶活力，在生理条件下，也会切除新合成DNA 5′端的RNA引物和填补所形成的缺口。这说明 Pol Ⅰ 在大肠杆菌的复制中也是有作用的。

大肠杆菌的3种DNA聚合酶虽然都能催化dNTP的聚合，但其特性并不相同(表5-2)，这反映它们在复制中所承担的作用不同。

表5-2　大肠杆菌DNA聚合酶Ⅰ，Ⅱ，Ⅲ

	Pol Ⅰ	Pol Ⅱ	Pol Ⅲ(核心)
结构基因	*polA*	*polB*	*polC*[a]
分子量(kD)	103	90	130,27.5,10[b]
亚基数	1	≥4	≥10
分子数/细胞，估计值	400		10～20
聚合酶活力	+	+	+[c]
3′→5′核酸外切酶活力	+	+	+
5′→3′核酸外切酶活力	+	−	−
模板-引物			
完整双股DNA	−	−	−
具引物长单股链	+	−	−
具缺口的双股DNA或5′端突出的单链			
单链长度<100核苷酸	+	+	+
单链长度>100核苷酸	+	−	−
转换数C，估计值	600	30	9 000

a. α亚基的基因；b. α,ε和θ亚基的大小；c. 聚合时(37℃)的核苷酸数/min/酶分子。

Pol Ⅲ 是大肠杆菌的复制酶，它在细胞中是以多亚基复合物的形态存在和发挥功能，这种复合物称为DNA聚合酶Ⅲ全酶(DNA polymerase Ⅲ holoenzyme)。DNA聚合酶Ⅲ全酶可以分为几个层次(表5-3)。Pol Ⅲ 全酶中，Pol Ⅲ 核心酶(core enzyme)与聚合作用直接有关，Pol Ⅲ 的亚基组成为αεθ，其中α为基因 *polC* 的产物，具有聚合酶活力。ε为基因 *dnaQ* 的产物，为3′

→5′核酸外切酶。α 和 ε 亚基的混合后会产生 1∶1 复合物，这时聚合酶活力就增加 2 倍，3′→5′核酸外切酶活力增加 50～100 倍，接近核心酶的水平。θ 亚基的功能还不清楚，可能与亚基间的结合有关。除了不能使具引物的单链 DNA 和具切口的 DNA 复制外，Pol Ⅲ 的催化活力与 Pol Ⅰ 相似。Pol Ⅲ 全酶中，核心酶以外亚基的功能现虽还不很清楚，但它们的参与提高了反应的速率和持续性(表 5-4)。这可能与有些亚基使核心酶与模板紧密结合有关，从而加强了作用的持续性。其中，β 亚基是最重要的持续性因子(processivity factor)，两个 β 亚基围绕 DNA 形成钳状结构，DNA 能滑动而难于脱落。

表 5-3　Pol Ⅲ 全酶的主要亚基及其组装层次

亚　基	分子量(kD)	基　因	组装层次
α	130	*dnaE*	Pol Ⅲ (α, ε, θ)
ε	27.5	*dnaQ*	
θ	10		Pol Ⅲ′ (α, ε, θ, τ)
τ*	71	*dnaX*	
γ*	47.5	*dnaX*	Pol Ⅲ* (α～ψ)
δ	35		全酶 (α～β)
δ′	33		
χ	15		γ 复合体 (γ～β)
ψ	12		
β	40.6	*dnaN*	

*γ 和 τ 亚基由同一基因所编码，而 γ 亚基由 τ 亚基的 N-端部分构成。

表 5-4　不同层次 Pol Ⅲ 的持续性

层　次	持续性	刺激因子
Pol Ⅲ	10	无
Pol Ⅲ′	60	亚精胺
Pol Ⅲ*	200	SSB
全酶	$>10^5$	SSB

Pol Ⅲ 全酶的结构是一种非对称的二聚体(asymmetric dimer)在图 5-14 中 Pol Ⅲ 包括 α，ε 和 θ，Pol Ⅲ′包含两个核心酶加上 τ 亚基，Pol Ⅲ* 比全酶少 β 亚基。括号中的数字代表分子量。这种结构特点支持了 DNA 在先导链和后随链上同时合成的看法(见本章第三节)。

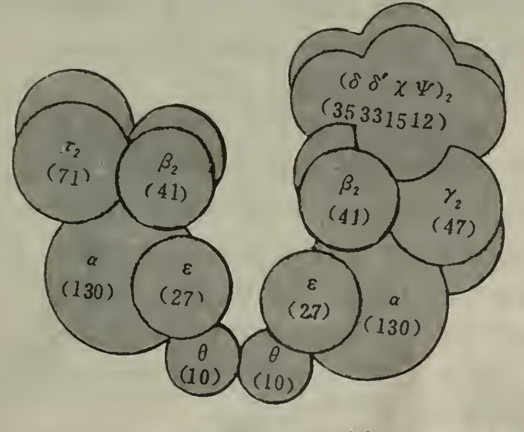


图 5-14　Pol Ⅲ 全酶的非对称二聚体的可能结构

二、引物的合成和引物酶

已经证明，DNA 的复制是半不连续的。既然聚合酶只能使 DNA 从引物的 3′端延伸，而不能引发 DNA 的重新合成，这就产生了另一问题，即 DNA 复制是如何开始的。噬菌体 M13 为此提供了一个线索，即它的 DNA 复制显示出对转录抑制剂的敏感性，从而表明 RNA 可能提供了 DNA 复制的 3′端。

Reiji 等利用含甲苯的缓冲液处理大肠杆菌以提高对外源核苷酸的通透性，然后再以大肠杆菌与 α-[^{32}P]脱氧核苷三磷酸和未标记的核糖核苷三磷酸的混合物保温，最后从大肠杆菌分

离 DNA 和用稀碱处理。碱性水解发生在 3′,5′-磷酸二酯键的磷酸基与 C-5′间，这样就使 dNTP 的放射性磷转移到核糖核苷酸上去(图 5-15)。结果发现，每一冈崎片段都产生一个 RNA-DNA 接头，说明 DNA 的复制是以 RNA 为引物。催化 RNA 引物合成的酶叫引物酶(primerase)，它是由大肠杆菌的 *dnaG* 基因所编码的。DNA 复制是随模板特定序列上 RNA 引物(表 5-5)的合成开始的。

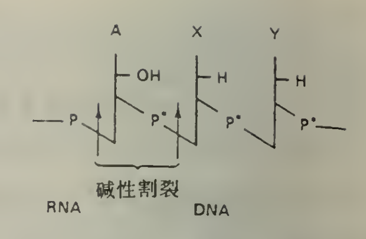


图 5-15　RNA 引物存在的证明

表 5-5　DNA 半不连续复制的引物

基因组	引物长度	主要产物
T7 噬菌体	1～5	pppApCpC/A$(pN)_{1\sim2}$* (N：多为 A 和 C)
T4 噬菌体	1～5	pppApC$(pN)_3$
大肠杆菌	1～3	未测定
*Φ*X174 噬菌体	1～5	未测定
海胆	1～8	(p)ppA/G$(pN)_7$
多瘤病毒、SV40 (分离核)	～10	pppA/G$(pN)_9$
动物细胞	～9	

* 在模板上合成的引物序列为 3′-pCpTpGpG/Tp-;G/T。

机体之所以选择 RNA 作为引物的原因可能在于减少致死突变(lethal mutation)。DNA 复制中，在模板上放置的最初几个核苷酸的正确性必然远比其后的为低。如引物为 RNA，这一部分所发生的差错就会避免或减少，因为引物随后即被降解。

三、DNA 双股链的分离和解螺旋酶

复制过程中，复制叉在不断前进，复制叉前方的亲代 DNA 就需不断解链，为使复制能够顺利进行，也就必须防止 DNA 单链的复性。以上过程是由解螺旋酶(helicase)和单链结合蛋白(single-strand binding protein，SSB)所承担(图 5-16)。

解螺旋酶利用 ATP 水解的能量使 DNA 双股链分离，并在 DNA 上沿一定的方向移动。生物机体含有多种解螺旋酶，有些解螺旋酶参与 DNA 复制，有些解螺旋酶参与 DNA 修复、重组等非复制过程。大肠杆菌的复制解螺旋酶有 DnaB、PriA 和 Rep 蛋白，它们的特性有些不同，移动的极性也有差异，DnaB 蛋白沿 5′→3′方向移动，PriA 和 Rep 蛋白沿 3′→5′方向移动。

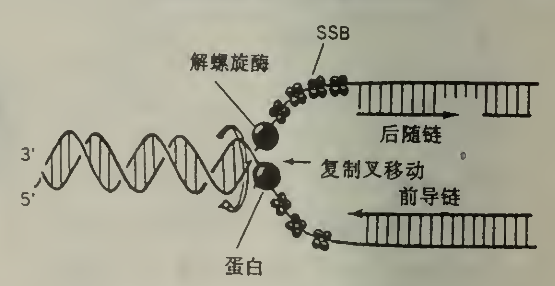


图 5-16　DNA 双螺旋的解旋

随着双螺旋 DNA 的逐步解旋，SSB 随之结合于单股链上。SSB 与单股 DNA 的结合是重要的，它们使后者保持一种伸展的构象，从而便于以之作为 DNA 合成的模板。SSB 对单链 DNA 的结合具有高度的协同性(cooperativity)，即 SSB 的初步结合能使其随后的结合更为容易。不同生物中，SSB 对单链 DNA 的结合机制似不相同。SSB 可以重复使用。SSB 还参与 DNA 的修复和重组。

四、超螺旋的松弛和旋转酶

天然 DNA 的负超螺旋虽然有利于 DNA 的解旋，但随着复制的进行，复制叉前方的亲代 DNA 中仍然会积累巨大的张力。如大肠杆菌以～1 000 核苷酸/s 的速率复制，这意味着 DNA

将以～100 rpm/s 的速率旋转，更何况大肠杆菌 DNA 是环型的，这种旋转实际上不可能，结果必然使复制停止。实际情况是Ⅱ型拓扑异构酶（旋转酶）参与了大肠杆菌的 DNA 复制，它不断解除 DNA 由此产生的拓扑学障碍。

拓扑异构酶在 DNA 复制中的关键性作用已通过抑制剂的应用得到证明。扑拓异构酶Ⅱ是由亚基 A 和亚基 B 所组成的四聚体，即 A_2B_2。萘啶酮酸（nalidixic acid）是一种复制抑制剂，但它的结合位点是拓扑异构酶的亚基 A。新生霉素（novobiocin）是另一个复制抑制剂，它结合于亚基 B 从而抑制与该亚基相联系的 ATP 酶活力。这些都证明拓扑异构酶的参与是 DNA 复制所必不可少的。

五、DNA 连接酶

DNA 后随链上的复制是不连续的。当冈崎片段形成后，其 5′端的 RNA 引物通过 DNA 聚合酶Ⅰ催化的切口移位而降解，并为脱氧核苷酸片段所置换，新形成的切口即由 DNA 连接酶（DNA ligase）予以封闭（图 5-17）。

图 5-17 DNA 连接酶的作用

DNA 连接酶在使双股 DNA 中的单链切口连接时，要求切口必须含 3′-羟基和 5′-磷酰基，并正确排列。连接酶利用 NAD^+ 或 ATP 中磷酸酐键的基团转移势以形成磷酸二酯键。反应过程中，先是 NAD^+ 或 ATP 的腺苷酰基转移给酶活性中心的赖氨酸使之活化，继而生成的中间体又以腺苷酰基转移给切口 5′端的磷酰基。切口 3′端羟基对磷酰基的亲核攻击就生成磷酸二酯键。真核生物和 T4 噬菌体利用以 ATP 作为腺苷酰体的供体，而原核生物则以 NAD^+ 为辅因子。

第三节　原核生物 DNA 的复制

DNA 复制，即使在原核生物中也是一个复杂的顺序性过程。除了 DNA 聚合酶外，它还需要许多辅助蛋白。这些酶和蛋白目前可能虽还不齐全，它们的功能也了解得不彻底；但各方面的证

据表明，这些酶和蛋白是协同作用的，是通过形成一种组织化的复合体以适应这一顺序性过程的。

DNA 复制的知识，不少是通过大肠杆菌的突变体获得的，也有不少是通过噬菌体获得的。所以，我们在进行原核生物 DNA 复制的讨论前，简单介绍这两方面的情况。

一、大肠杆菌与复制有关的基因及突变体

DNA 复制涉及到许多酶和有关蛋白。这些知识很多来自大肠杆菌的相应突变体，它们在确定有关酶和蛋白如何开展工作的研究中起过重要作用。例如，人们发现有些基因如 *dnaA* 控制新复制叉的形成，但对已存在复制叉的移动无影响；另一些基因，则由于编码聚合反应组分（如 *dnaE* 编码 Pol Ⅲ 的 α 亚基）或引物体组分（如编码引物酶的 *dnaG*）参与已存在复制叉的移动。又如编码 Pol Ⅰ 的 *polA* 基因和 DNA 连接酶的 *lig* 基因的突变体亦已发现。这些位点的突变不会使 DNA 合成停止，但引起冈崎片段的异常积累。这些情况说明了各个蛋白在 DNA 复制中所承担的角色和所起的作用。现在，许多涉及 DNA 合成的基因已经克隆（clone），已获得便于研究的数量，这些进展对 DNA 复制的研究无疑是重要的。

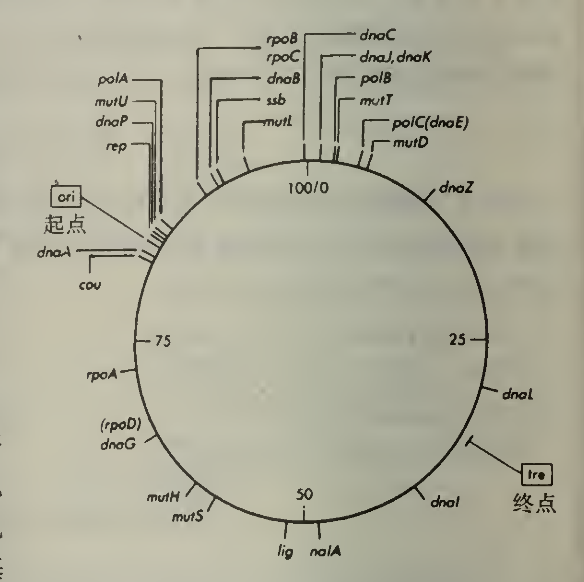


图 5-18　与 DNA 复制部分有关的大肠杆菌 K 12的遗传图

表 5-6　大肠杆菌复制基因

基　因	图中的定位(min)	突变体的表型	蛋白质及其体外功能
dnaA	82	慢停，缺陷性原点起始	未知
dnaB	91	快停	DnaB 蛋白，引物预合成
dnaC	99	慢或快停（与突变有关）	DnaC 和 DnaB 蛋白复合
dnaE(*polC*)	4	快停	聚合酶 Ⅲ 全酶 α 亚基
dnaG	66	快停，缺陷性片段起始	引物酶
dnaI	39	慢停	未知
dnaJ	0.5	慢停	未知
dnaK	0.5	慢停	未知
dnaL	28	快停	未知
danP	84	苯乙醇抗性，慢停	未知
dnaT	95～99	调节终止	未知
dnaZ	10	快停	聚合酶 Ⅲ 全酶的 γ 亚基，链的生长
cou	82	香豆霉素（和新生霉素）抗性	拓扑异构酶 Ⅱ 的亚基 B，切割/封闭
lig	51	复制片段的积累	DNA 连接酶
nalA	48	萘啶酮酸（和奥啉酸）抗性	拓扑异构酶 Ⅱ 的亚基 A，ATP 酶活力
polA	85	DNA 修复能力受损	DNA 聚合酶 Ⅰ
rep	83	复制叉缓慢移动	解螺旋酶
ssb	91	快停	SSB

大肠杆菌的许多有关DNA复制的基因现在已定位(图5-18),它们的产物功能也已清楚(表5-6)。此外,枯草杆菌和酵母一些有关基因的定位也已完成。

二、噬菌体 ΦX174 的复制

噬菌体是DNA复制研究中的一个重要实验系统,这是由于它们的基因组都较小,它们的复制是在进入寄主细胞后主要利用寄主的基因产物进行的。噬菌体DNA有的是线型,有的是环型;有的是单链,有的是双链。这些特点都对复制研究提供了方便,下面就先讨论ΦX174的复制。

大肠杆菌噬菌体ΦX174的基因组是一个含5 386核苷酸的单链环型DNA,复制中首先合成的不是其自身而是其互补链。为了进行繁殖,ΦX174采取的策略是:先合成原来DNA(称为正链)的互补链(称为负链),然后(+)链和(-)链形成双链环型DNA,这种双链环型DNA称为复制型(replicative form,RF),然后以RF中的(-)链为模板合成(+)链,即原来的DNA。或是说,ΦX174的复制循环可以分为3个阶段:(1)病毒(-)链的合成和RF的形成(SS→RF);(2)RF的繁殖(RF→RF);(3)新病毒(+)链的产生(RF→SS)。这一过程中的(-)链合成已成为研究大肠杆菌中后随链复制的模型,而由复制型生成(+)链则成了先导链上DNA合成的模型。

ΦX174的以(+)链为模板的(-)链合成(SS→RF)完全依靠寄主的酶和有关蛋白,这一过程需要引物体的参与。引物体(primosome)由7种蛋白组成(表5-7)。ΦX174的DNA在接近2 300位处有一个由44核苷酸组成的发夹结构,它与引物体的形成有关。整个过程可分为6步(图5-19)。

表5-7 引物体的组成

蛋 白	亚基结构	亚基质量(kD)
PriA(n′)	单体	76
PriB(n)	二聚体	11.5
PriC(n″)	单体	23
DnaB	六聚体	50
DnaC	单体	29
引物酶(DnaG)	单体	60

n,n′,n″为各该蛋白原来的表示法。

(1)当ΦX174的(+)链进入寄主细胞后,除了分子中的发夹结构外,其它部分都为SSB所覆盖。引物体的组装是以PriA蛋白对引物体组装位点(primosome assembly site,pas)的识别开始的,ΦX174的pas位于基因*F*和*G*间的70核苷酸片段上,pas能形成包括发夹结构在内的多种二级结构而且能被识别。随后,PriB和PriC参与复合物的形成。

(2)由PriA,PriB和PriC组成的复合物进一步与DnaT,DnaB和DnaC组装成预引物体(preprimosome)。其中DnaB和DnaC形成由ATP稳定化的B_6-C_6复合物。预引物体与引物酶的结合形成了引物体。

(3)PriA利用ATP水解的能量使引物体在(+)链上沿3′→5′方向移动。

(4)引物体在随机选定的位置逆转移动方向,并在多个位置合成RNA引物。DnaB是一种依赖于DNA的ATP酶,它驱动引物体在ssDNA上的移动,但方向与PriA相反。

(5)Pol Ⅲ全酶从引物延伸合成DNA片段。

(6)Pol Ⅰ切除RNA引物,填补缺口,连接DNA片段。在这一过程中所形成的带切口的双链DNA称为RFⅡ,切口封闭和超螺旋化后的复制型称为RFⅠ。

引物体继续与DNA相结合以备参与(+)链的合成。

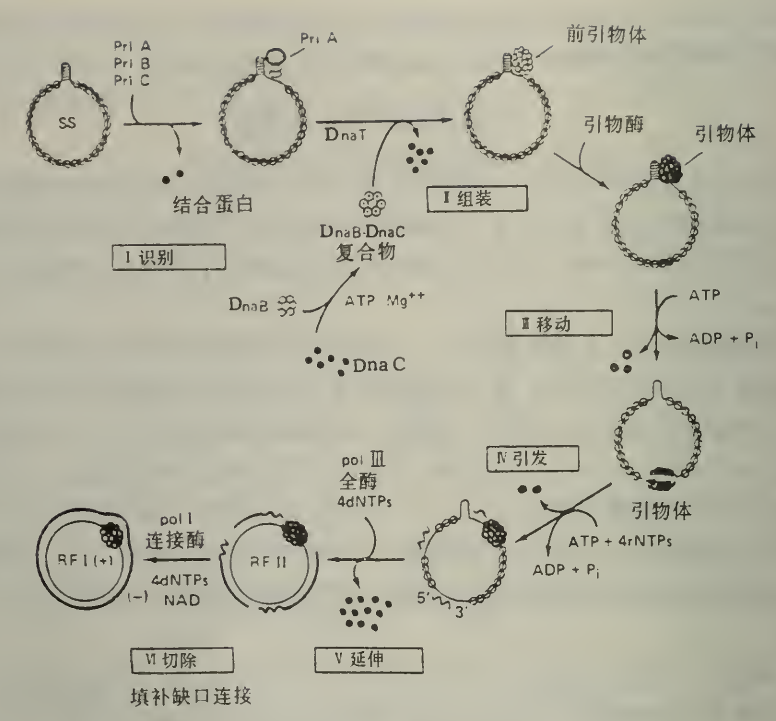


图 5-19　$\Phi X174$ 的以(＋)链为模板的 RF 合成

$\Phi X174$ 的由 RF 合成(＋)链的过程(RF→SS)是研究先导链复制的模型。图 5-20 表示出两种过程：一是不连续的(－)链合成；二是连续的(＋)链合成。(＋)链合成涉及到两种未参与 SS→RF 的蛋白，其中之一是噬菌体编码的基因 *A* 蛋白(gpA)；另一是大肠杆菌的 Rep 蛋白。

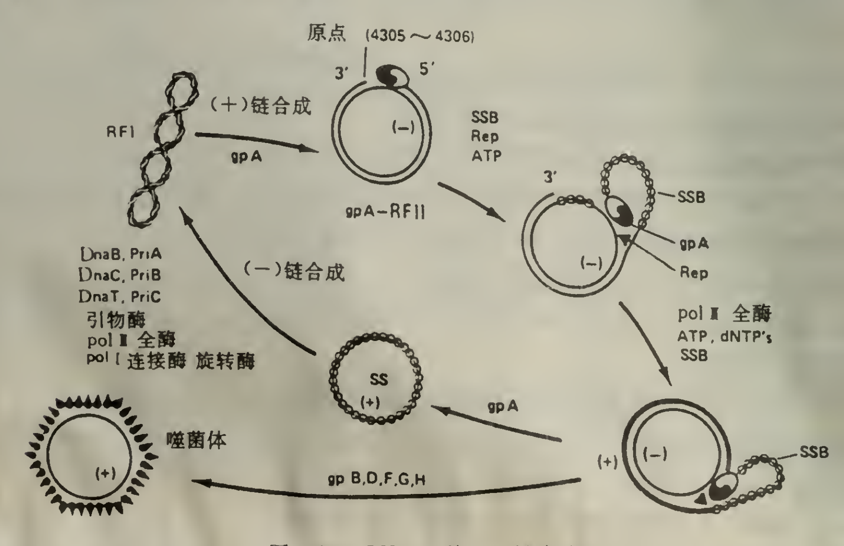


图 5-20　ΦX 174 的(＋)链合成

gpA 是一 60 kD、对 $\Phi X174$(＋)链原点具有专一性的核酸内切酶活力的蛋白，它是这一过程中 $\Phi X174$ 所编码的唯一复制蛋白。gpA 的功能是引发复制起始，整个(＋)链合成过程可以分为 4 步：

(1)(＋)链合成是从 RF Ⅰ 开始的。gpA 借助于引物体结合于识别位点，专一性地作用于4 305和4 306间的磷酸二酯键产生一切口。切口的 3′-OH 端为4 305位的 G 残基，蛋白则通过其酪氨酸残基结合于4 306位 5′-P 的 A 残基，从而保存键能。

(2)Rep 蛋白结合于 gpA 处的(－)链。Rep 蛋白从(＋)链使 gpA 复合物从(＋)链的 5′端开始使双股链不断解开，从双股链中分离出来的(＋)链即为 SSB 覆盖，以免重新与(－)链退火。

(3)Pol Ⅲ 全酶从原来(＋)链的 3′端使 DNA 链不断延伸，结果就形成一个环化的滚环结构(looped rolling circle structure)。其中原来的(＋)链在复制叉移动时仍然与 gpA 相连接，这犹如原来的(＋)链逐步被剥离 RF。

(4)复制绕(－)链超过完整一周后即产生一双股复制原点，gpA 在此再做专一性切割，并与新形成的 5′端共价结合。gpA 的第二次切割生成一单位长度的 *ΦX*174 DNA，它的 3′-OH 向 gpA 5′-P 做亲核攻击从而形成环状分子。每一分子的 gpA 含两个活性酪氨酸，故可完成相继的切割，并与 5′端连接。

在 *ΦX*174 感染的中期，每一新合成的(＋)链指导(－)链合成从而形成 RF。在感染的后期，新合成的(＋)链通过噬菌体编码的病毒成熟蛋白和衣壳蛋白(gpB,D,F,G 和 H)形成新的病毒粒子。

由上述过程可见，*ΦX* 174 的(－)链合成是不连续的，相当于大肠杆菌 DNA 后随链的复制；(＋)链的合成是连续的，相当于大肠杆菌先导链的复制。

三、大肠杆菌的 DNA 复制

大肠杆菌的基因组是一双股环型 DNA。它的复制是由单一原点出发按双向的 θ 方式进行的。先是原点处双股 DNA 解链，然后是先导链上 DNA 的连续合成和后随链上 DNA 的不连续合成，最后复制完成。

（一）复制的起始

原核生物DNA 由于较小，每一分子只有一个复制原点，而真核生物DNA 则有多个原点。DNA 复制原点不同于后随链上冈崎片段的起始点。大肠杆菌基因组的复制原点位于天冬酰胺合成酶和 ATP 合成酶操纵子间，共有 245 bp，称为 *oriC*。现在已分离到许多噬菌体、细菌、质粒和细胞器的复制原点，并测定了它们的核苷酸序列，它们一般都含两个系列的重复单位（图 5-21）：3 个 13 bp 重复序列和 4 个 9 bp 重复序列。大肠杆菌编码的 Dna 蛋白即作用于 *oriC* 而引发复制的起始，Kornberg A. 认为大肠杆菌的起始经历了下述步骤（图 5-22）：首先是 DnaA 蛋白识别和结合于 *oriC* 的 9 bp 重复单位形成复合物，其中 20～40 个 DnaA 蛋白形成核心，负超螺旋的 *oriC* 包裹于外。HU 蛋白（参阅第四章）参与并促进这一过程。随后，DnaA 蛋白使 13 bp 重复单位熔解而形成开放复合物。Pl 核酸酶（作用于单链）的作用证明约有 45 bp 成为单链。开放复合物的形成需要 DnaA 蛋白和 ATP。这一过程发生于 22℃以上（至少在离体条件下如此）。于是 DnaA 蛋白使 DnaB 和 DnaC 蛋白进入 DNA 的熔解区形成前引物化复合物（prepriming complex）。DnaB 为解螺旋酶，它使 DNA 双向熔解形成两个复制叉，并使引物酶和 RNA 聚合酶得以进入，这两种酶参与了先导链上引物的合成。在这一过程中，RNA 聚合物的参与称为转录活化（transcriptional activation），此外，SSB 和旋转酶也都参与了这一过程。

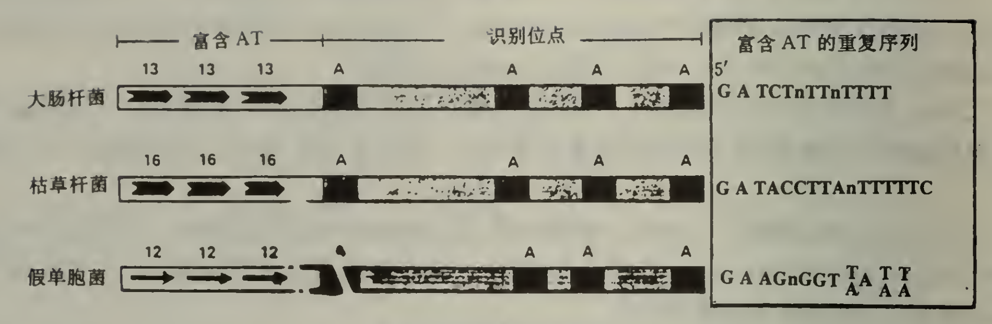


图 5-21　细菌的复制原点

A 为构成 DnaA 蛋白识别位点的 9 bp 重复单位；箭头表示富含 AT 的重复单位；n 代表任何核苷酸。

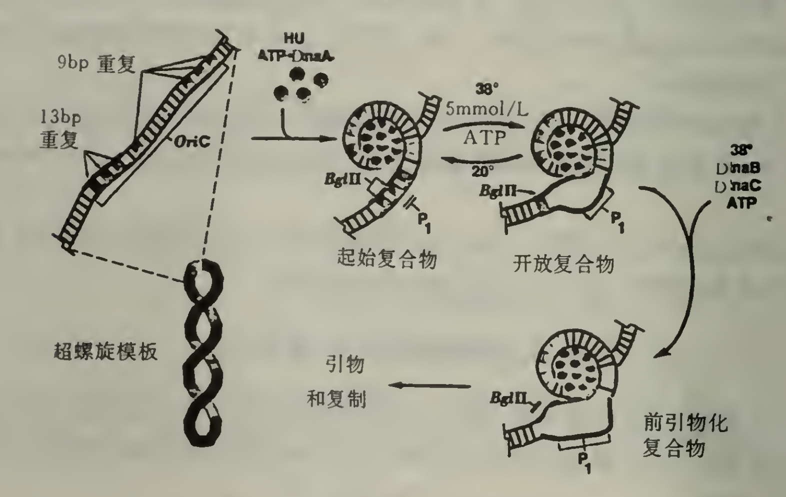


图 5-22　大肠杆菌 DNA 复制的起始

大肠杆菌中，DNA 复制是受细胞分裂所控制的，每一细胞分裂周期仅复制一次，37℃时，大肠杆菌的加倍时间随生长条件而在＜20 min～10 h 之间变化。不过，复制叉的移动速率是固定的，每秒约 850 核苷酸，这就使复制时间固定在 40 min，加上分裂过程中细胞组分的分离和隔膜的形成约需 20 min，两者相加约共需 60 min。因此，加倍时间少于 60 min 的细胞，在前一轮细胞周期结束前必已开始新一轮的复制，这已为实验所证实。从而表明，细胞中可能存在一种触发每一轮 DNA 复制的信号。这种调控机制目前还不清楚，但各方面的研究已指出可能的机制之一是：DnaA 蛋白由与其紧密结合的 ATP 的缓慢水解（约 1 h）而形成钝化的 Dna A-ADP 复合物。这一复合物的活化是通过以 ATP 置换 ADP 实现的，而细菌质膜中的酸性磷脂能促进这一过程。复合物由活化向钝化状态的转化，以及一种称为染色体起始抑制剂（inhibitor of chromosomal initiation，Ici A）的蛋白结合于 *oriC* 的 13 bp 重复单位，就阻止了不适当时候复制的起始。这一过程的细节现在还不清楚。

（二）复制的延伸

随着复制起始中 θ 结构的形成，许多酶和蛋白相继进入复制叉（图 5-23）。来自噬菌体、动物病毒和 *oriC* 研究的知识说明，这些酶和相关蛋白在复制叉形成了称为复制体（replisome）的超分子复合物。不同系统中，复制体的组分虽有差异，但其功能十分相似。解螺旋酶使 DNA 双

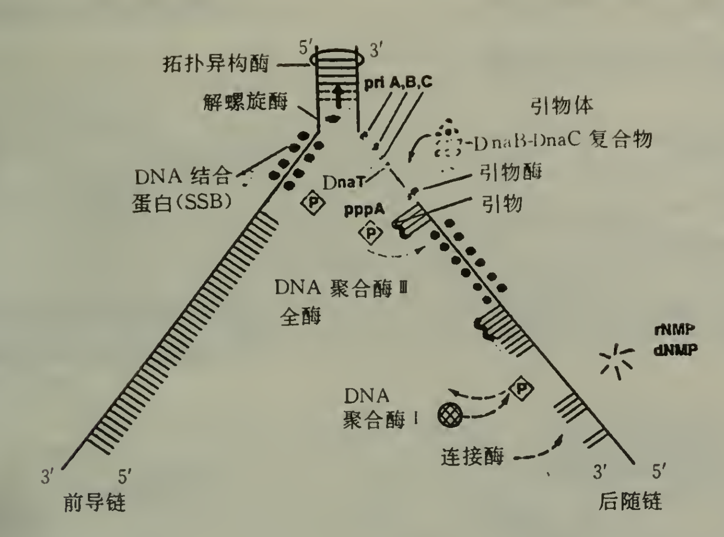


图 5-23　DNA 复制叉

股链分离和复制叉前进；旋转酶解除由此产生的正超螺旋张力；DNA 单股随即为 SSB 覆盖，形成作为模板所需的伸展构象；前导链上开始 DNA 的连续合成，后随链上的不连续合成；冈崎片段上引物的切除、缺口的填补和切口的封闭。根据计算，每一复制叉仅含 1 个 DNA 聚合酶 Ⅲ 全酶，这说明前导链和后随链上的 DNA 合成是由同一复制体负责的。但 DNA 两股链上的 DNA 合成不是同步进行的，并且方向相反。那么，复制体是怎样工作的呢？对此，现在已提出了一个模型。这一模型认为，当前导链上开始 DNA 合成和复制叉向前移动时，后随链随即向后回折成环，并与聚合酶的另一个活性中心按前导链的取向缔合（图 5-24）。在这样的结构中，PriA 和 DnaB 蛋白是两个关键的组分，PriA 蛋白具依赖于 DNA 的 ATP 酶活力，它提供引物体移动的动力，清除过程中遇到的 SSB，使 DnaB 蛋白和引物酶发挥功能。DnaB 蛋白的解螺旋活力在大肠杆菌的复制中占重要地位，它使复制叉处的 DNA 不断解旋。但这两种蛋白的移动极性不同，DnaB 蛋白为 $5' \rightarrow 3'$，Pri A 为 $3' \rightarrow 5'$。这一模型可以合理地解释在 DNA 的双股链上是如何同时合成 DNA 的，使引物酶和复制叉的相反移位方向得到统一解释。复制时，DNA 不断地在复制体中抽动，不断使 DNA 延伸。

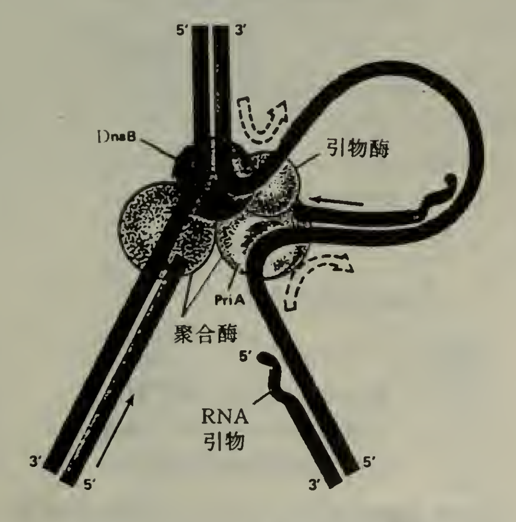


图 5-24　复制叉处前导链和后随链同时复制的工作模型

（三）复制的终止

复制的终止（termination）是一个尚不清楚的问题，大肠杆菌的基因组是一环型 DNA，它的复制终止位点约相对于原点（见图 5-18）。我们知道大肠杆菌 DNA 复制进行到一特定区域

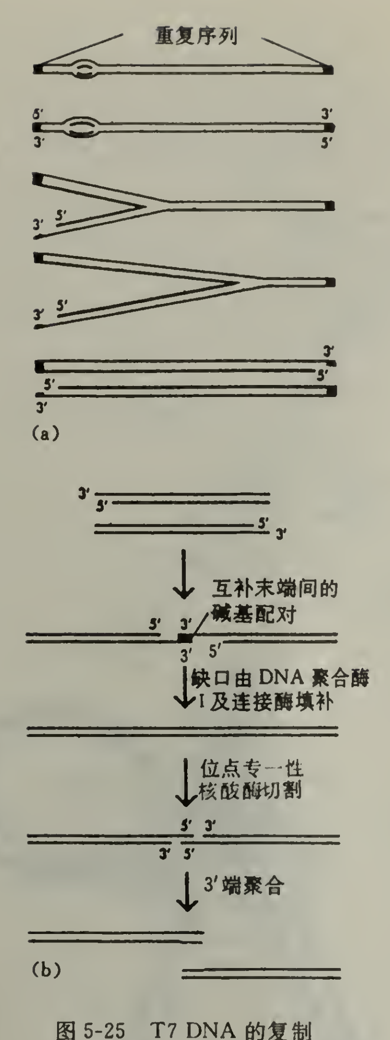


图 5-25　T7 DNA 的复制

(a)线型 DNA 3'端的不完全复制；

(b)多联体的形成和复制的完成。

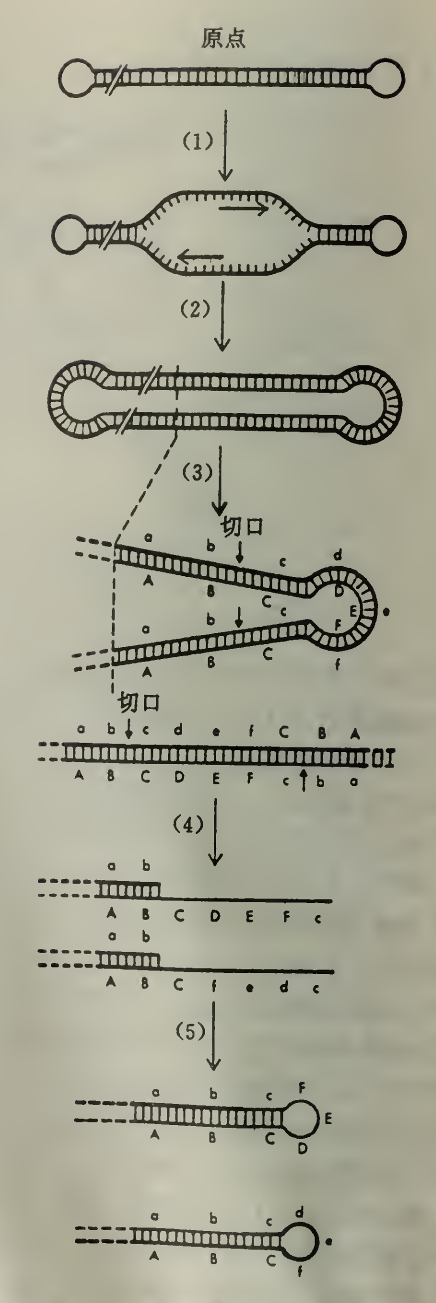


图 5-26　痘病病毒的 DNA 复制

图中由(3)到(5)仅表示出半分子。

后，复制叉的移动会由于这一区域两侧的核苷酸序列而放慢移动速度，从而协调两复制叉到达的时间。在复制的最后阶段，大肠杆菌的环形 DNA 会产生两个相互套接在一起的环。那么，大肠杆菌 DNA 的复制究竟最后怎样完成并生成两个独立的 DNA 分子的呢？拓扑异构酶在这里显然发挥了重要作用。当复制接近完成时，两个复制叉相互接近所产生的空间障碍会干扰拓扑异构酶对复制叉前方 DNA 的解旋作用。这时，一个可能的情况是，DNA 先通过拓扑异构酶的作用先去连环化，然后 DNA 再完成复制。

虽然很多细胞、病毒和质粒的 DNA 是环型的，但也有一些基因组是线型的。这些线型 DNA 的复制，由于 RNA 引物在复制后的切除，故而在 5'端会留下一缺口。这些线型 DNA 是

怎样完成其末端的复制的呢? T7 DNA 的复制(图 5-25)提供了一种途径。

T 7 DNA 的复制原点位于分子左端约 17%处,复制双向进行,不久左端的复制完成,形成一 Y 结构并不断扩大。由于 T7 DNA 的两个末端是重复的,其左端完全是其右端的重复。故 T 7 DNA 复制生成的不是单位长度的子代DNA,而是由单位长度DNA 通过其末端重复序列连接成的多联体(concatemer)。多联体为单位长度 DNA 间提供了 DNA 延伸所需的 3′端,然后 Pol Ⅰ 填补缺口,连接酶封闭切口,多联体再经专一性核酸酶切割成单位长度分子。

痘病病毒(poxvirus)为线型DNA 末端复制提供了另一途径(图 5-26)。痘病病毒的两端由发夹结构连接。复制从其内部的起始位点开始(1),复制使发夹结构转变成双链(2),分子内部发生交错切割(3),分子重排(4),末端回折再经 DNA 连接酶作用生成原来的分子(5)。

原核生物线型 DNA 复制的终止现仍不完全清楚。

(四)复制的忠实性

DNA 复制是一个高度精确的过程。据计算,每复制 10^8~10^{10} bp 中仅有一个错配。复制的这种近乎完全的高度精确性是必要的,因为只有这样才能保证遗传信息在传代中保持完整。

复制的忠实性是通过多种因素实现的。聚合酶 Ⅰ 和 Ⅲ 的 3′→5′核酸外切酶活力对此做出了贡献。现在,聚合酶 Ⅰ 的结构已经确定,其Klenow 片段具有两个结构域。大结构域含一个荷正电的、直径约 2 nm 的裂缝,DNA 即结合于此。DNA 一旦结合,裂缝即被关闭;这时,DNA 只能在裂缝中前后滑动。而小结构域含核苷酸结合位点。两结构域在空间相距 3 nm。当新合成的 DNA 中含错误配对的核苷酸时,双螺旋发生变形,因而不再能在裂缝中向前滑动,只能后移,这时 3′→5′核酸外切酶活力就切除错配的核苷酸。错配核苷酸切除后,聚合反应继续进行(图 5-27)。聚合酶中聚合和校对两种活力的相对强弱影响着突变的频率。如聚合酶对 3′-5′核酸外切酶相对活力高,突变就上升,反之就下降。

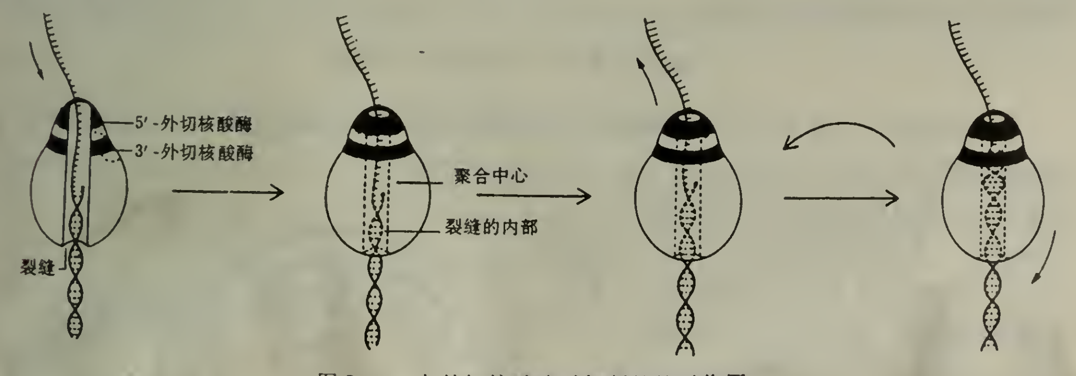


图 5-27　3′-外切核酸酶对复制的校对作用

DNA 聚合酶只能从引物的 3′端延伸 DNA 链,也增加了复制的精确性。这可以从两方面理解:碱基配对具有协同性,开始几个碱基错配的概率高;同样,开始阶段所形成的短核苷酸序列中的错配也不易被校正。由于 RNA 引物最终被降解,其中可能存在的错误也就为随后由聚合酶 Ⅰ 合成的正确短 DNA 片段所代替。

后随链上的 DNA 合成是不连续的,但生物机体为何未演化出一种能使核苷酸按 3′→5′方向聚合的酶,从而使后随链上的 DNA 合成也是连续的呢? 实际上这一系统有利于提高复制的忠实性。如核苷酸按照 3′→5′方向聚合、延长中的 DNA 链 5′端必须保留三磷酸基以驱动下一步反应(图 5-28)。如聚合反应出现差错,错配碱基切除后留下的只能是 5′-OH 或 5′-P。这些

末端基团都难于为进一步的聚合提供所需的能量。为使聚合反应得以继续，5′端必须重新活化。因此，不连续合成虽然复杂，但它有利于错配核苷酸的校正。

图 5-28　假想的脱氧核苷酸的 3′→5′ 聚合反应

复制虽然具有校正功能，但总还可能有漏校。此外，环境中的化学或物理因素还不断地使 DNA 受到损伤，它们可以通过细胞的各种修复机制不断加以纠正，这就把 DNA 中的差错降低到最小程度。

第四节　真核生物 DNA 的复制

许多证据说明，真核和原核生物 DNA 复制的基本特征是相同的。但由于真核生物 DNA 远大于原核的，大多数真核生物是由不同的细胞组成的多细胞系统，故而两者间尤其在复制的调节上显示出许多明显的不同。

一、真核生物 DNA 复制的一些特点

真核生物的一生要经历许多次细胞分裂，而每一次细胞分裂都要经历 4 个不同的时期，称为细胞周期(cell cycle)。细胞周期可分为：①有丝分裂期(M)；②第一间隔期(G1)，决定细胞进入 DNA 复制或是停留在静止期(G0)，如细胞一旦通过限制点(restriction point)，它就进入 S 期；③合成期(S)，它是细胞周期中唯一发生 DNA 合成的时期；④第二间隔期(G2)，细胞随后进入有丝分裂(图 5-29)。

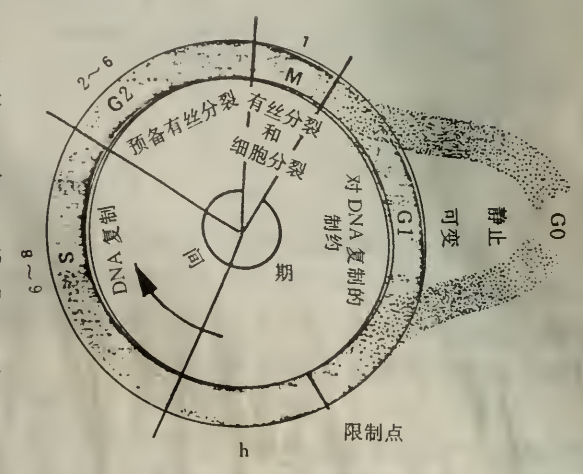


图 5-29　真核生物的细胞周期

培养细胞的细胞周期通常可经历 16～24 h。在多细胞生物中，不同类型细胞的周期可相差很大，从 8 h 直至 100 天以上，这种差异多发生在 G1 期。有些细胞，如神经元等则已不再分裂。

在细胞周期中，细胞能否通过限制点决定它进入细胞分裂或是滞留在静止期，

这种信号来自细胞外。对多细胞生物而言，它是通过一些专一性的生长因子结合于细胞表面的受体发挥作用的，不同类型细胞具有对应于不同生长因子的受体。例如，上皮细胞的繁殖是由来自血小板的生长因子(PDGF)引起的，T 细胞是由白细胞介素 2 激发的。生长因子与受体的结合就触发有关的级联系统(cascade system)。

真核和原核生物的另一个显著不同点是，前者的 DNA 具有多重复制原点，而原核生物只有一个。在真核生物中，复制叉的移动速率为约 50 核苷酸/s，如从一个原点出发进行双向复制，一个平均大小的真核基因组的复制就需要 1 个月时间。这与事实不符，细胞周期中的 S 期一般只经历几个小时。电镜观察证明，真核生物染色体具有多重原点，约 3～300 kb 一个原点。真核生物 DNA 上的原点数决定于物种和组织两者。

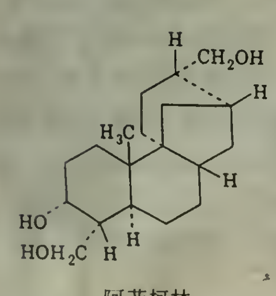
真核生物 DNA 原点的特征，由于其基因组的巨大和检测技术上的困难，现在还不清楚。但有证据表明，酵母 DNA 中存在专一性的复制原点，它们称为自主性复制序列(autonomous replication sequence，ARS)。酵母约含 400 ARS，每一复制子约为 36 kb。

ARS 是由两个功能域组成的：(1)功能域 A 含 11 bp，其共有序列为(A/T)TTTAT(A/G)TTT(A/T)。这是 ARS 的核心，可能是起始蛋白的结合位点，不过这有待确证；(2)功能域 B 富含 AT，延伸到核心序列 3′-端 50～100 bp 处，这可能是 DNA 熔解的区域。ARS 显然是可能的复制原点，但仍有许多问题不清楚。

实验证据表明，真核生物 DNA 的各个区域不是全都同时复制的，而是 20～80 相邻复制子一次被活化。S 期中不断有新复制子被活化，直至整个染色体完全复制。真核生物病毒的 DNA 合成研究说明，复制子的复制是由引物转录触发，由与原点相联系的增强子与转录因子的结合所调控的。组织特异和细胞周期特异的转录因子能够使特定的复制子进入复制，从而控制每一细胞周期仅复制一次。

DNA 复制是细胞分裂的中心问题，DNA 在何时和以何种频率进行复制是受多种信号控制的。

二、真核生物 DNA 聚合酶



阿非柯林

人们对真核生物 DNA 聚合酶比原核生物的酶了解得更少，这主要由于它们分离纯化不易，不同组织和不同时期细胞中的酶有差异，难于得到相应的突变体等。迄今在真核细胞的核中已发现 4 种 DNA 聚合酶：α，β，δ 和 ε；另一种 γ 存在于线粒体内(表 5-8)。

DNA 聚合酶 α 在体内参与染色体 DNA 的复制。它的这一功能是通过抑制剂阿非柯林(aphicolin)确定的。阿非柯林是一种真菌产物，它在抑制聚合酶 α 活力的同时抑制了 DNA 复制，这证明它参与了复制。不同真核细胞 Pol α 的结构和性质都相似，其核心亚基具聚合酶活力，50 kD 和 60 kD 的亚基具引物酶活力，但未检测到 3′→5′ 核酸外切酶活力。

Pol δ 与 α 一样，都参与 DNA 的复制，不同的是，Pol δ 缺乏引物酶活力，但显示 3′→5′ 核酸外切酶活力。它们的另一个不同是，Pol δ 具高持续性，而 Pol α 的持续性仅为中等，约 100 核苷酸。这些特点表明：聚合酶 δ 可能是前导链的复制酶，仅偶尔需要引物，但要求高的持续性；聚合酶 α 则是后随链的复制酶，它的 100～200 核苷酸的持续性已可满足冈崎片段的合成，但需经常形成引物。聚合酶 α 和 δ 可能是整个复制体的两个组分。

表 5-8 真核生物 DNA 聚合酶

	α	δ	ε	β	γ
细胞内的部位	核	核	核	核	线粒体
分子量(kD)					
天然状态	7 250	170	256	36～38	160～300
核心酶	165～180	125	215	36～38	125
其它亚基	70,50,60	48	55	无	35,47
性质					
对 PCNA[a] 的应答	−	+	−	−	−
持续性	低	高[b]	高	低	高
保真性	高	高	高	低	高
复制	+	+	+	−	+
3′→5′ 核酸外切酶	−[c]	+	+	−	+
引物酶	+	−	−	−	−
修复	−			+	−

a. PCNA 代表增殖细胞核抗原;b. PCNA 存在时;c. 果蝇中为隐性。

聚合酶 δ 具有多种形式。由牛胸腺得到的聚合酶 δ 由 125 kD 和 48 kD 两种亚基组成;酵母的聚合酶 δ 由 125 kD 和 55 kD 亚基组成;由人胎盘分离到的相应酶含 1 个 170 kD 多肽和一些较小的亚基。Polδ 的一个重要特征是对增殖细胞核抗原(proliferating cell nuclear antigen,PCNA)的依赖性。PCNA 大量存在于增殖细胞的核中。由牛胸腺分离得到的 PCNA 为一 36 kD 的蛋白质。PCNA 使 Polδ 的持续性增加 40 倍,它的作用类似 Pol Ⅲ 全酶中的 β 亚基。

聚合酶 β 和 ε 的功能尚不完全清楚,可能与 DNA 修复有关。

真核生物 DNA 复制还涉及两种称为复制因子(replication factor,RF)的蛋白,即 RFA 和 RFC。它们存在于从酵母到哺乳动物中。RFA 为真核生物单链 DNA-结合蛋白,相当于原核生物的 SSB。RFC 促进活性复制复合物的组装。

三、真核生物 Pol γ 和线粒体 DNA 复制

线粒体是真核生物的重要细胞器,它有自身的 DNA。线粒体 DNA 一般为环型双股分子。动物的线粒体 DNA 的大小约 16 kb,植物的线粒体 DNA 更为复杂。线粒体 DNA 的双股链由于浮力密度的不同而区分为轻链和重链。线粒体 DNA 的复制(图 5-30)称为 D-环复制,它含有两个单向复制叉,负责这一复制的是真核生物 DNA 聚合酶 γ。复制从重链(H 链)的原点开始,这时新合成的 H 链即置换原来的链,这样所形成的结构称为取代环(displacement loop)或 D 环。当 H 链的合成进行到约 2/3 时,轻链(L 链)合成的原点即被暴露,从而引发了 L 链的合成,后者延伸的方向与 H 链相反。整个过程约需 1 h,这和真核生物 DNA 的复制远慢于原核生

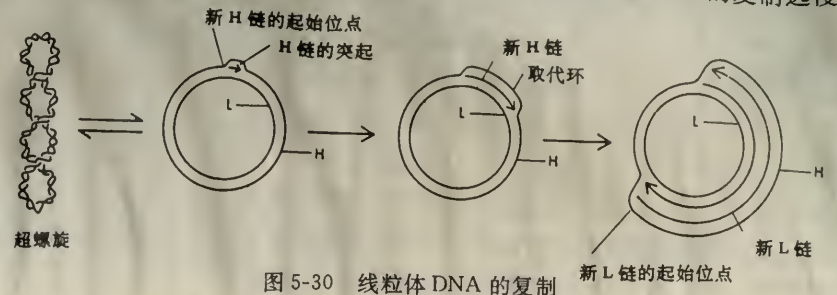


图 5-30 线粒体 DNA 的复制

物 DNA 相一致。这一复制方式大致亦需 RNA 引物。人们亦在线粒体中发现了拓扑异构酶，它们可能参与亲链的解旋和子链的分离。

四、逆转录酶

逆转录酶的存在是在 70 年代证实的。一些含 RNA 的肿瘤病毒如鸟类成髓细胞白血病病毒(avian myeloblastosis virus)和劳氏肉瘤病毒(Rous sarcoma virus)含有一种独特的 RNA 指导的 DNA 聚合酶(RNA-directed DNA polymerase)，它以 RNA 为模板指导 DNA 合成。感染时，病毒的单链 RNA 连同逆转录酶一起进入寄主细胞，逆转录酶即以病毒 RNA 为模板合成与其互补的 DNA 链，同一酶使 RNA-DNA 杂交分子中的 DNA 为模板合成互补 DNA 从而形成双股 DNA。这样合成的双股 DNA 被整合进一寄主基因组。在一定条件下，整合的休眠病毒基因组被活化，随即转录生成病毒 RNA 和形成新的病毒粒子。这一过程可以表述为：

$$\mathrm{RNA}(+) \longrightarrow \frac{\mathrm{RNA}(+)}{\mathrm{DNA}(-)} \longrightarrow \frac{\mathrm{DNA}(-)}{\mathrm{DNA}(+)} \longrightarrow \mathrm{RNA}(+)$$

逆转录酶除具有 RNA 指导的 DNA 聚合酶活力外，还具有：核糖核酸外切酶(核糖核酸酶 H)活力，它专门降解 RNA-DNA 杂交分子中的 RNA 和 DNA 指导的 DNA 聚合酶活力。

含逆转录酶的 RNA 病毒也称为逆转录病毒(retrovirus)。逆转录病毒基因组含 3 种基因(图 5-31)：*gag*，*pol* 和 *env*。*gag*(group associated antigen)基因编码一种多蛋白(polyprotein)，它通过自身的蛋白水解酶活力裂解为 3～4 种病毒核心蛋白。*pol* 编码逆转录酶。逆转录酶有 α 和 β 两个亚基；*pol* 编码的为 β 亚基(分子量 90 kD)，而 α 亚基(分子量 65 kD)是 β 亚基的水解产物。*pol* 产物还能使蛋白水解生成使病毒 DNA 整合进寄主基因组所需整合酶(integrase)。*env* 编码另一多蛋白，它产生构成被膜的蛋白。在病毒 RNA 的两端为长末端重复(long terminal repeat，LTR)，它长几百个核苷酸，与病毒基因的调节有关。

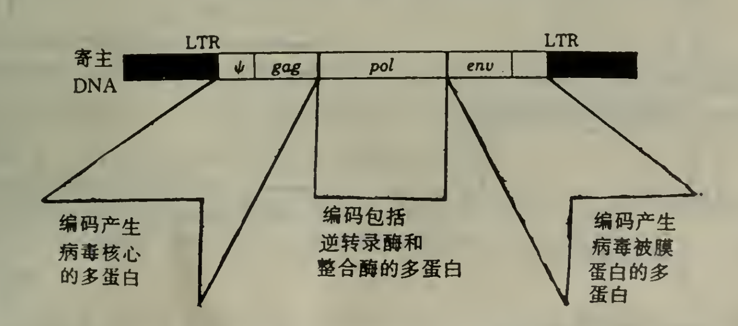


图 5-31 整合后的逆转录病毒基因组

ψ 是还原病毒 RNA 包装进成熟病毒粒子所需的序列。

逆转录酶与其它 DNA 和 RNA 聚合酶一样含 Zn^{2+}。逆转录酶作用时也需引物，它是包含于病毒粒子内的一种 tRNA，它与病毒 RNA 5′端的互补序列相配对从而提供 DNA 延伸所需的 3′端，DNA 合成同样按 5′→3′方向进行。逆转录酶与 RNA 聚合酶一样不具 3′→5′核酸外切酶活力，聚合的差错率达 1/20 000 核苷酸，其后果是容易突变。

逆转录酶在理论和应用上都有重要意义。逆转录病毒引起癌症和艾滋病。逆转录酶也与真核生物基因组中一些重复序列的来源有关。真核的许多转座子具有与逆转录病毒基因组十

分相似的结构，在其两翼为LTR，编码区包含有还原病毒逆转录酶同源的序列，它们有时亦称为逆转录转座子(retrotransposon)。它在基因组中可能通过RNA中间体的逆转录实现转座。逆转录酶在遗传工程中也很有用，它可以使mRNA，tRNA等转录成其互补DNA(complementary DNA，cDNA)并进行基因克隆。

五、DNA端粒的复制

真核生物线型DNA的末端为具特定结构的端粒(telomere)(见第四章)。它们一般含许多串联的短寡聚核苷酸序列，其一股链如为T_xG_y，其互补链就为C_yA_x，x和y在1～4范围内。

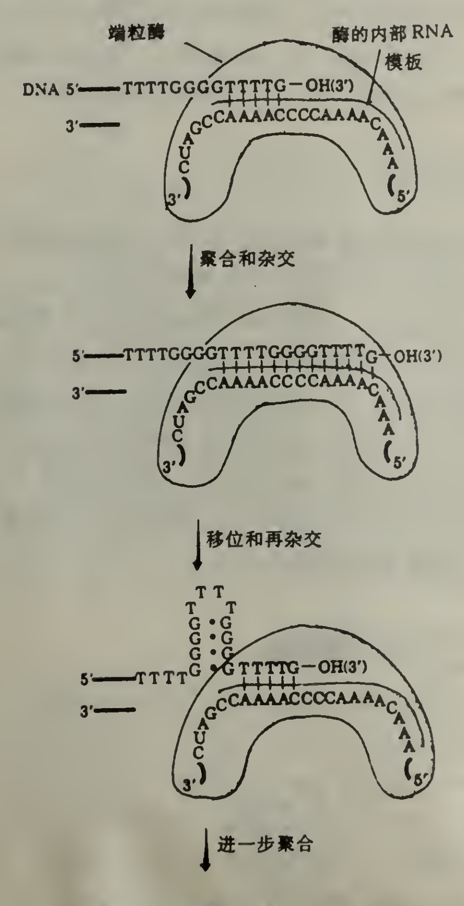


图5-32 端粒合成的一种模型

端粒结构具有特定的生物学意义。DNA复制需要引物，但在线型DNA分子末端不可能通过正常的机制在引物被降解后合成相应的DNA片段(参阅原核生物线型DNA分子末端的复制)。如无特殊的机制合成末端序列，染色体就会在细胞传代中变得越来越短。这一难题是通过端粒酶(telomerase)的发现才得到了澄清。端粒酶能使端粒延伸，它是一种含RNA的蛋白复合体。端粒酶所含的RNA约长150核苷酸，并约含1～5拷贝的C_yA_x重复序列，它是合成端粒T_xG_y股的模板。端粒酶实际上是一种逆转录酶，它催化互补于RNA模板的DNA片段的合成。

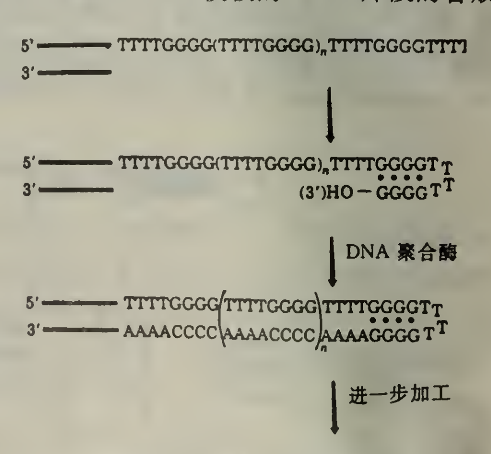


图5-33 端粒合成的完成

端粒酶的作用，在已讨论过的酶中是找不到先例的，它的机制也不很清楚，但已提出了一些模型，如图5-32。图中表示的是四膜虫的端粒合成。酶的内部模板先是通过碱基配对与端粒的3'端的TG股结合，然后根据模板的序列延伸TG股。在合成一拷贝的重复序列后，端粒酶必须调整位置以便进一步延伸端粒。这是通过新合成的TG股回折并藉非标准碱基配对形成发夹结构实现的，这称做尺蠖模型。当TG股进一步延伸至一定长度时，又回折形成发夹结构，从而使其末端成为合成互补CA股所需的引物(图5-33)。

四膜虫中，端粒酶活力的消失使端粒长度逐步缩短，最后导致细胞死亡。类似情况亦存在人体中，生殖细胞的端粒长度维持不变，而体细胞的端粒长度则随个体的老化而逐步缩短。对此的一个推论是，生殖细胞具端粒酶活力，体细胞则否。这一问题的解决无疑会有助于对生命衰老的认识。

第五节　DNA 的修复

DNA 复制虽然是一个高保真的过程，但复制形成的子代 DNA 总不免还会存在少量未被校正的差错。除此之外，DNA 还会受到各种物理和化学因素的损害。这些差错和损伤如不予以改正，结果就会引起突变，甚至机体的死亡。

一、DNA 损伤的来源

环境中的许多因素都可引起 DNA 结构上发生某些变化。DNA 的损伤归纳起来主要有：

（一）碱基的脱落

酸和热都能影响糖苷键而使碱基脱落。在生理温度和 pH 下，哺乳动物的基因组据估计每天约失去 10 000 嘌呤和几百嘧啶碱基。许多由其它因素所产生的已改变的碱基也会为专一性的 DNA 糖基化酶所除去，从而形成无嘌呤或无嘧啶位点或 AP 位点(apurinic or apyrimidinic site)。

（二）碱基(或核苷)的改变

电离辐射和亲电试剂(如烷化剂)等都能改变碱基和糖基。硫酸二甲酯、甲烷磺酸甲酯(MMS)等都是常见的烷化剂，它们能生成相应的正碳离子与碱基中的亲核基团作用生成烷基

$$(CH_3O)_2SO_2 \qquad CH_3-\overset{O}{\underset{O}{\overset{\|}{\underset{\|}{S}}}}-OCH_3 \quad \text{MMS}$$

化碱基。碱基类似物(如 5-溴尿嘧啶为胸腺嘧啶的类似物)的掺入 DNA 也使原来的碱基转换成另一碱基。黄曲霉毒素 B1，2-乙酰-氨基芴、苯并芘等都是强力的致癌物，能在 DNA 中导入大体积的加合物，从而在复制过程中改变遗传信息。

由于自身的化学不稳定性或亚硝酸的作用，胞嘧啶、腺嘌呤和鸟嘌呤能自发形成尿嘧碇、次黄嘌呤和黄嘌呤，其结果是尿嘧啶与腺嘌呤配对，黄嘌呤与胞嘧啶配对。

（三）错误碱基

其复制中错误碱基的掺入会导致形成错配碱基对。增变基因(mutator gene)编码具缺陷的聚合酶、核酸酶或高保真复制所需的其它蛋白，故也提高了突变频率。碱基的互变异构也能导致复制中错误碱基的掺入。

（四）碱基的缺失或插入

二氨基吖啶（NH_2 取代的三环结构）

吖啶类(acridine)由于其分子呈扁平，易于嵌入 DNA 碱基平面间(图 5-34)，从而导致在复制或重组过程中缺失或插入一个碱基。此外，DNA 聚合酶在复制过程中发生滑移，尤其在连续存在几个相同碱基的区段产生 1 个或几个碱基的缺失或插入。聚合酶在模板链上的滑移易于造成缺失，在生长链上的滑移易于造成插入(图 5-35)。

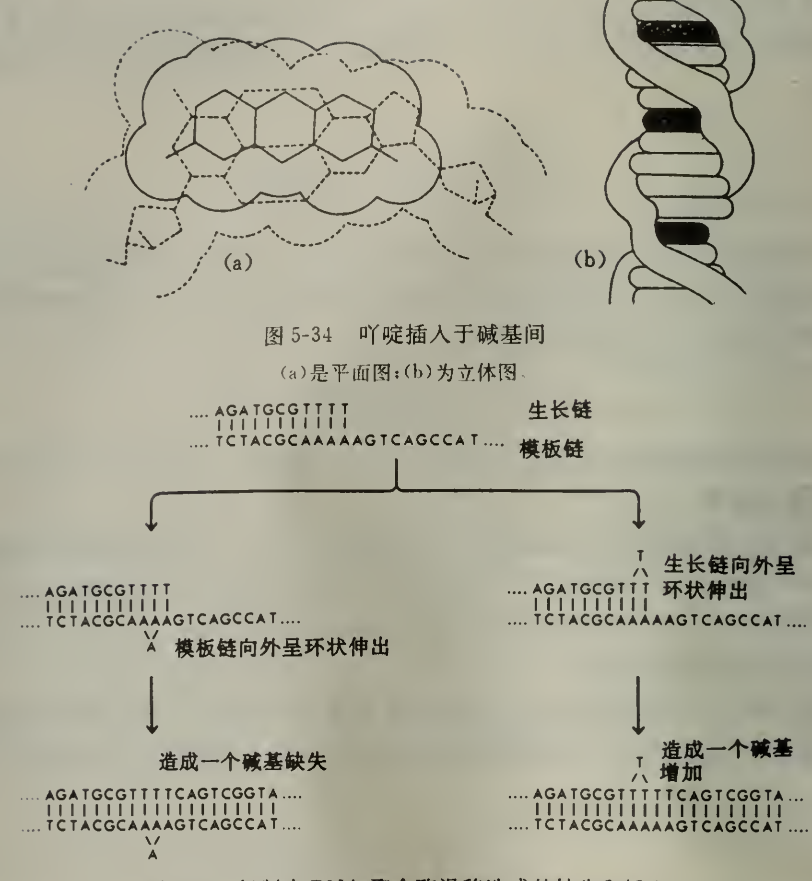


图 5-34　吖啶插入于碱基间

(a)是平面图；(b)为立体图

图 5-35　复制中 DNA 聚合酶滑移造成的缺失和插入

（五）嘧啶碱基的二聚化

紫外线照射(254 nm)能使相邻嘧啶、尤其是胸腺嘧啶形成环丁基二聚体。胞嘧啶(或胸腺嘧啶)的 C_6 也可与相邻胞嘧啶的 C_4 连结形成另一重要的光化学产物(图 5-36)。这些二聚体的存在就使正常的碱基配对难于发生。

（六）链的断裂

电离辐射或博莱霉素(bleomycin)等化学试剂能作用于 DNA 链，使单股或双股链的磷酸二酯键发生断裂。

（七）3′-脱氧核糖片段

许多能产生游离基的试剂(博莱霉素和电离辐射)都能使脱氧核糖裂解，结果就在链的 3′端产生磷酸乙醇酸这样的裂解产物。

（八）DNA 链的交联

二价烃化剂，丝裂霉素(mitomycin)和光活化的补骨脂(psoralen)等能够使 DNA 双股链间形成交联，其结果使双股链不再能分开。

在引起 DNA 损伤的各种原因中，氧化是普遍、严重而又了解得很少的一种。在原核和真核细胞中，γ 射线和紫外线照射，臭氧、过氧化物都能通过各种形式产生不同的效果。在这一过

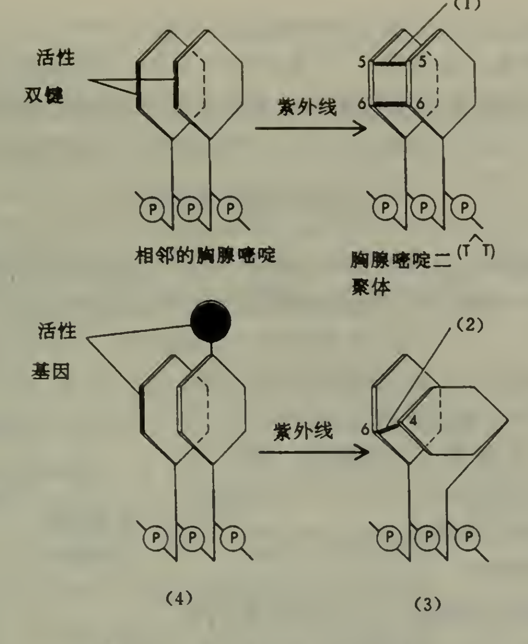


图 5-36　紫外线照射产生的嘧啶

(1)每个嘧啶环的第 5 和第 6 个碳跨环连接;(2)5′嘧啶的 6 位碳与 3′嘧啶的 4 位碳相连;(3)胸腺嘧啶-胞嘧啶 6～4 产物;(4)相邻的胸腺嘧啶(左)和胞嘧啶(右)。●为 NH_2。

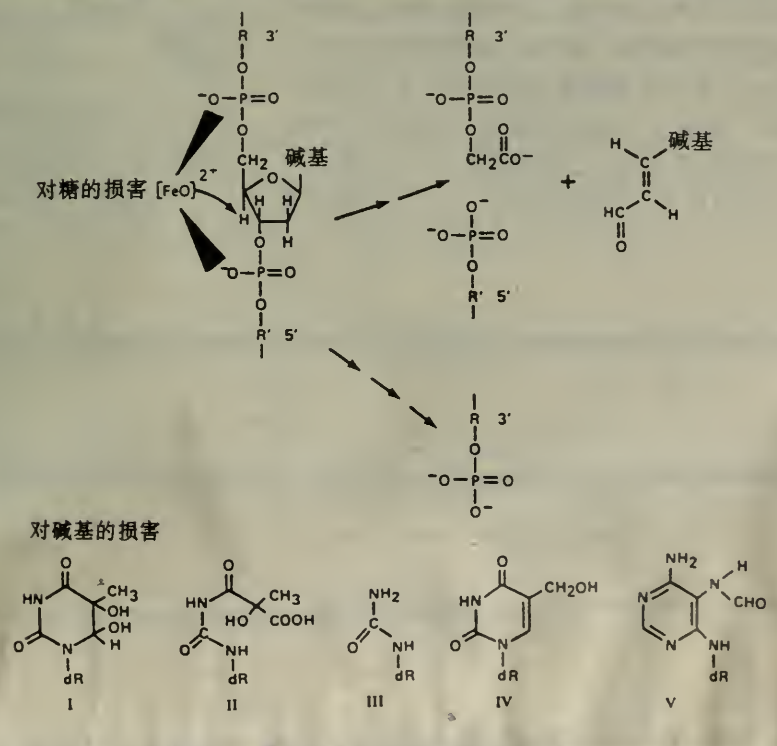


图 5-37　氧自由基生成的 DNA 损伤产物

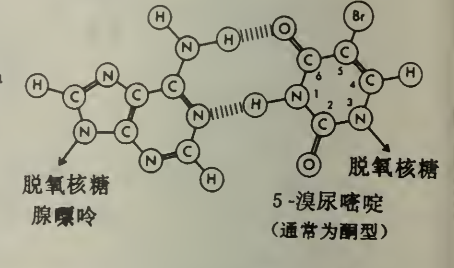
程中，氧逐步被还原为水：$O_2 \longrightarrow O_2^{\cdot} \longrightarrow H_2O_2 \longrightarrow HO^{\cdot} \longrightarrow H_2O$
氧自由基对DNA 的损伤是由金属离子，尤其是铁离子所介导的(图 5-37)。因此，螯合剂、自由基清除剂和过氧化物歧化酶、过氧化物酶和过氧化氢酶活力的增强都能降低氧的毒性。

$$Fe^{2+} + H_2O_2 + H^+ \longrightarrow Fe^{3+} + HO^{\cdot} + H_2O$$

二、DNA 修复与突变

虽然DNA 聚合酶的校对功能和DNA 的修复机制十分有效，但复制产生的一些差错仍可能会漏校，以及DNA 的一些损伤仍可能未被修复，结果它们就会遗传下去。DNA 的这种永久性的改变叫做突变。

突变可由单一碱基的改变产生，称为点突变。这种一个碱基的改变可以是一个碱基的缺失(deletion)、插入(insertion)，或是由一个碱基由一种嘧啶（嘌呤)转变为另一种嘧啶(嘌呤)的转换(transition)，或是由嘧啶（嘌呤)转换为嘌呤(嘧啶)的颠换(transversion)。突变也可由多个碱基的改变产生。



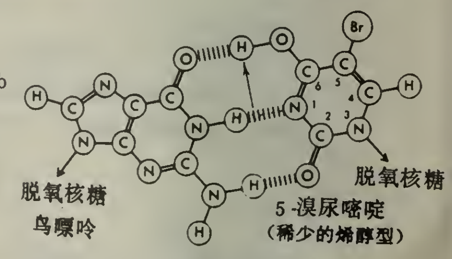


图 5-38　5-溴尿嘧啶的两种配对方式

DNA 复制中的错配以及随后产生的差错都可能形成突变。现在我们举例讨论这一过程。

前述的碱基的烷基化、自发脱氨基或是由于亚硝酸的作用或是由于碱基类似物的掺入都可以在一定程度上改变核苷酸序列，即改变遗传信息。5-溴尿嘧啶是胸腺嘧啶的类似物，当它的酮型存在时能与腺嘌呤配对(A－BU)(图 5-38 a)，当它以烯醇型存在时能与鸟嘌呤配对(G－BU)(图 5-38 b)。这是基团键合性质改变的结果，当5-BU 掺入到DNA 中时，它就可能使 A－T 转换为G－C(图 5-39)。但在某一情况下 5-BU 再次掺入 DNA 并与G 配对(呈烯醇型)，G－C 又可能转换为 A－T，这称为回复(reversion)。

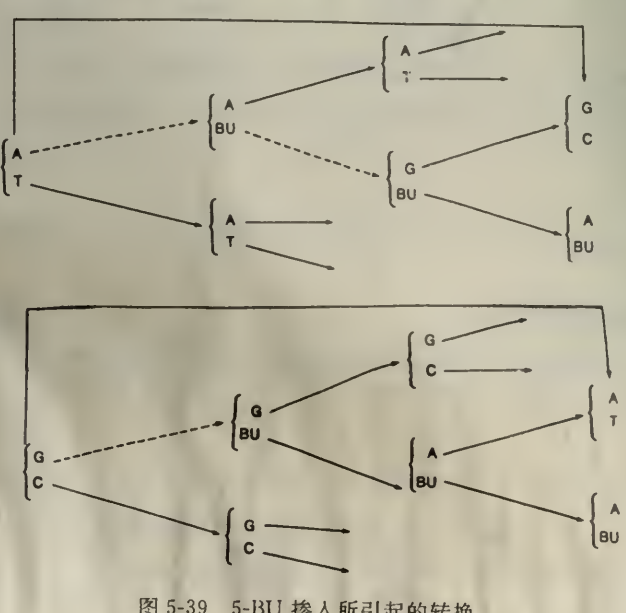


图 5-39　5-BU 掺入所引起的转换

颠换的机制还不清楚。现已发现，烷化剂能够导致脱嘌呤作用，嘌呤一旦除去，4 种碱基可能进入该位置，从而可能导致颠换。

点突变仅改变基因的一个密码子。由于遗传密码具有简并性，故这一改变可能或并未引起一种氨基酸为另一种氨基酸所替代，如密码子 GGT 改变为 GGA，两者都编码脯氨酸，故不会引起产物组成和结构上的改变。即使产物中个别氨基酸确发生了变换，但如两种氨基酸的性质和结构相近，产物的生物学特性仍可无明显改变。这种突变常称为沉默突变(silent mutation)。不过，有些改变如影响到关键性氨基酸，其结果就会导致产物生物功能的丧失。假如这一酶又位于中心代谢途径中，这种突变就会致死，称为致死突变(lethal mutation)。

还有一种可能的情况是，某一氨基酸改变的结果是基因产物的功能基本上未有重大变化，改变的只是 K_m 和最大速度等，这种突变称为渗漏突变(leaky mutation)。有时，基因突变产物的功能可能更适合机体所处的特定条件，从而更有利于其存活，这种突变就可能导致新物种的出现。

当基因突变是由于碱基的缺失或插入引起时，其结果是 DNA 的密码子与所生成的多肽链中的氨基酸残基间的共线性(colinearity)就会被中断。这种中断从缺失和插入部位开始，随后发生读码框(reading frame)的移动。在mRNA 中，每一个核苷酸三联体编码一种氨基酸，称为密码子。在原则上，mRNA 上的三联体序列可按 3 种可能方式翻译(图 5-40)。实际上只有其中之一产生功能性蛋白，这种密码子序列称为读码框，碱基的缺失或插入常引起读码框的移动，故称为移码突变(frameshift mutation)(图 5-41)。大多数移码突变常产生无活性产物，有些移码突变会在基因内部形成终止密码子，从而使所生成的多肽链长度变短。

5′ 3′
1 CUC AGC GUU ACC AU
—Leu—Ser—Val—Thr—
C UCA GCG UUA CCA U
—Ser—Ala—Leu—Pro—
CU CAG CGU UAC CAU
—Gln—Arg—Tyr—His—

图 5-40 蛋白质合成中的 3 种可能的读码框

移码突变有时可由同一基因内的另一碱基的缺失或插入所抵消，但要求第二个移码突变发生在前一突变位点的下游。这样的第二个移码突变称为校正突变(suppressor mutation)。

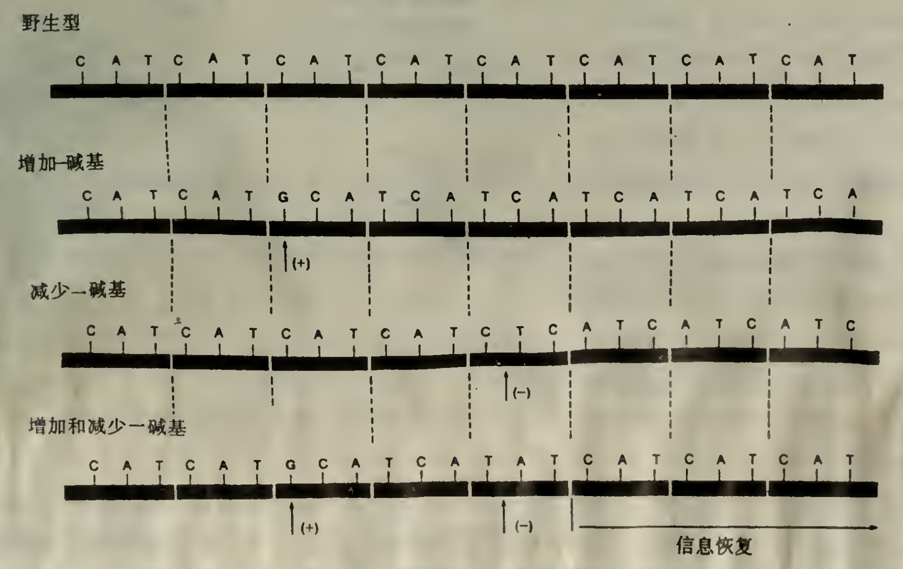


图 5-41 移码突变和校正突变

三、DNA 修复的多样性

DNA 是细胞中唯一的一种在受损伤或改变后能修复的分子，因为没有其它任何分子完整性的受损会像 DNA 那样影响细胞的存活。机体的各种基因必须保证高度的稳定，DNA 之所以能表现出这种稳定性的原因之一即在细胞具有多种修复机制。

DNA 的修复机制是多样(图 5-42)而有效的。大肠杆菌基因组约有 100 个位点涉及 DNA 的修复和有关功能。DNA 的互补双链保证它一股链上的损伤被切除后能从另一股获得所需的遗传信息。同时，核苷酸的缺失和插入，链间的交联也可通过重组予以置换。一个在复制过程中被漏校的不适当碱基也可通过错配修复(mismatch repair)予以改正。甚至损伤严重的双股链的断裂也可通过连接酶封闭和通过重组予以修复。

复制
单链损伤
易错复制
原位修复
无差错切除修复
子链缺口
易错修复
无差错
(重组)修复
股间交联
双链断裂

图 5-42　DNA 的多种修复机制

DNA 的修复机制可以分为：①使损伤逆转，如光活化、脱烷基化等；②切除受损部分，然后通过复制或重组予以置换；③错配修复。当修复无效时，基因组的连续性(continuity)可以通过修复得到保持，即复制可以通过损伤的旁路进行。

四、直 接 修 复

DNA 中的嘧啶二聚体可通过可见光(300～600 nm)的光解作用而恢复为单体(图 5-43)。催化这一过程的酶为光敏裂合酶(photolyase)。反应的第一步是酶结合于 DNA 中嘧啶二聚体部位。酶含两个发色团，一是 $FADH_2$；另一在大肠杆菌和酵母中为蝶呤，在其它生物中为脱氮黄素(deazaflavin)。当变形的 DNA 吸收与发色团特征性波长相当的光波后，酶-DNA 复合物就利用这能量使二聚体裂解为单体，酶随即脱离 DNA。

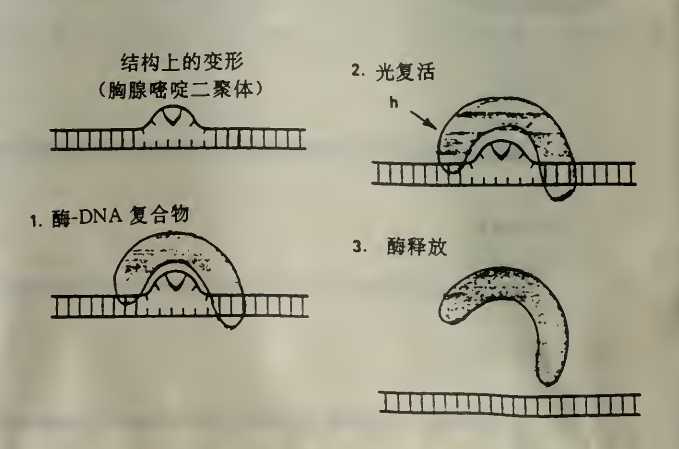


图 5-43　嘧啶二聚体的酶促光解

烷基化碱基也可通过类似的途径逆转。例如，o^6-甲基鸟嘌呤的甲基在 o^6-甲基鸟嘌呤-DNA 转甲基酶的作用下转移给酶分子的一个半胱氨酸残基。与一般的酶不同，该酶的这一过程是不可逆的。

五、切除修复

DNA 中的嘧啶二聚体也可通过切除而修复。在大肠杆菌中，担当切除功能的酶为多亚基的 UvrABC 核酸内切酶，它们是 *uvrA*，*uvrB* 和 *uvrC* 基因的产物。该酶作用时需要 ATP，它切割于二聚体 5′和 3′端的第八和第四磷酸二酯键，形成的缺口为 Pol Ⅰ 和 DNA 连接酶填补和封闭的。

切除修复中的核酸内切酶可切除嘧啶二聚体，也可切除其它受损伤部位（图 5-44），如：①受损伤的碱基，形成 AP 位点；②含 AP 位点的片段；③受损伤的核苷酸及邻近部位；④双链的交联部位。切除所造成的单链缺口为聚合酶（大肠杆菌中为 Pol Ⅰ）、DNA 连接酶填补和封闭的。

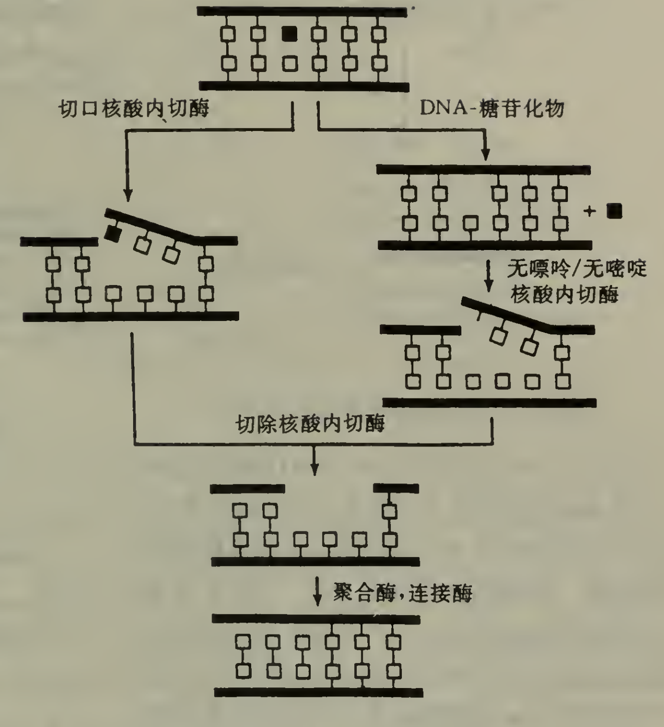


图 5-44　切除修复的两种途径

DNA 受伤部分的切除涉及许多核酸酶，如切口核酸内切酶，AP 位点核酸内切酶，切除核酸外切酶，UvrABC 和 N-糖基化酶等。真核生物的 DNA 切除修复现在还不清楚。

六、错配修复

复制是一个高保真过程，但其正确性毕竟不是绝对的，复制产物中仍会存在少数未被校出的错配碱基。错配碱基的修复会使复制的保真性提高 10^2～10^3 倍。现已在大肠杆菌、酵母和哺乳动物中都发现了这一系统。

错配修复是按模板的遗传信息来修复错配碱基的，因此，修复时首先要区别模板链和新合成的 DNA 链。这是通过碱基的甲基化实现的。

大肠杆菌 DNA 的 5′-GATC 序列中 A 的 N^6 都是甲基化的。这是 Dam 甲基化酶（专门使脱氧腺苷甲基化）负责的。复制后的一个短暂（约几分钟）的时间内，新合成链的 GATC 中的 A 未被甲基化，故子代 DNA 暂时是半甲基化的（图 5-45），几分钟后新合成链进一步甲基化，从而成为全甲基化 DNA。半甲基化 DNA 成为识别模板链和新合成链的基础。错配修复发生在 GATC 的邻近处，故这种修复也称为甲基指导的错配修复（methyl-directed mismatch repair）。

大肠杆菌的错配修复机制还不完全清楚，但所需的 Dam 甲基化酶，MutH，MutL，MutS 蛋白，DNA 解螺旋酶Ⅱ，SSB，DNA 聚合酶Ⅲ，核酸外切酶和 DNA 连接酶都已得到纯化，并重建了它们的工作模型（图 5-46）。在修复过程中，MutS，MutH 和 MutL 是关键性的，MutS 结合于错配碱基，MutH 结合于GATC，MutL 使 MutS 和 MutH 连结成复合物。如错配碱基位于距

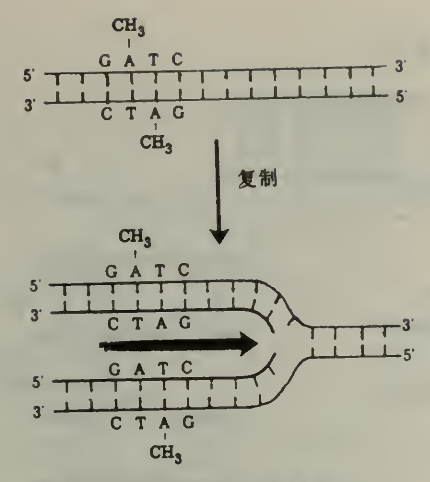


图 5-45 新合成的半甲基化 DNA

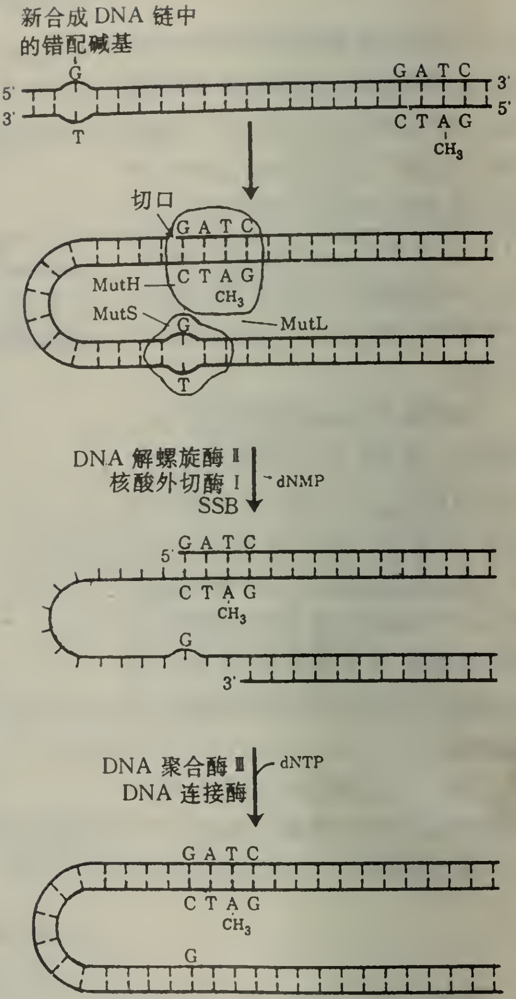


图 5-46 甲基指导的错配修复

GA^{m}TC 1 000 核苷酸以内，MutH 的位点专一核酸内切酶活力，就在未甲基化链中 GATC 序列中 G 的 5′侧切割，作为待修复链的标志。如错配位于切割点的 5′侧，未甲基化链即"解链者"按 3′→5′方向降解直至错配碱基在内，所缺少的部分由新链填补。这一过程需要 DNA 解螺旋酶 Ⅱ，SSB，核酸外切酶 Ⅰ（按 3′→5′方向降解单链 DNA），DNA 聚合酶 Ⅲ 和 DNA 连接酶的联合作用。如错配位于切割点的 3′侧，除了以核酸外切酶 Ⅶ（按 5′→3′或 3′→5′方向降解单链）或 RecJ 蛋白（一种按 5′→3′方向降解单链的核酸外切酶）代替核酸外切酶 Ⅰ 外，修复过程都相同。

所有错配都可由这一系统修复，但其中以 G－T 错配修复更为有效，C－C 错配的修复为弱。

由于要降解和置换 1 000 或更多碱基对，这种修复代价显然是高的，但这也说明修复的重要，细胞为此不惜代价。

七、易错修复

前述的各种修复都以受损伤链的互补链为模板进行的，因此是无差错的(error free)。但当 DNA 所受的损伤是严重和广泛时，修复常易招致差错(error prone)。

正常情况下，DNA 的损伤多数在复制前即已修复，但也还可能有少数未被修复。遇到这种情况，机体可能产生两种措施。一种措施是复制越过损伤部位，在其下游约 1 000 核苷酸处重新开始(图 5-47)。新合成链的缺少部分则通过重组修复(recombinational repair)予以补充，即以受损伤链的互补链的相应片段转移至缺口处，这时DNA链上仍然存在的损伤随后再为直

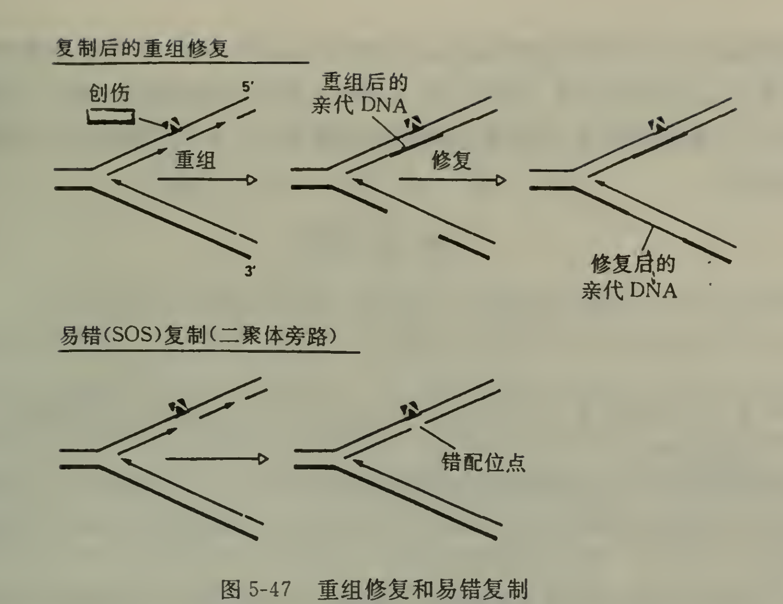


图 5-47　重组修复和易错复制

接修复或切除修复系统予以修复。机体对 DNA 的严重创伤可采取的另一种措施是产生一系列的蛋白来应急，故称为应急措施(SOS response)。这些应急蛋白的功能(表 5-9)有些已知，有些未知。其中一些蛋白负责超越 DNA 创伤复制(translesion replication)。这是一种无模板指导的复制(non-template directed replication)，因此，常易错配和致突变。负责超越创伤复制的聚合酶是聚合酶Ⅱ。DNA 聚合酶Ⅱ与聚合酶Ⅲ不同，它能够使复制通过 AP 位点这样的 DNA 损伤。它的有些亚基与聚合酶Ⅲ相同，有些亚基是在应急措施中大量生成的。除了聚合酶Ⅱ外，易错修复还需要其它的蛋白。

表 5-9　应急蛋白的功能和相应基因

基　因	在 DNA 修复中的作用
已知功能的基因	
polB(*dinA*)	编码易错修复所需的 DNA 聚合酶Ⅱ的聚合亚基
uvrA、*uvrB*、*uvrC*	编码 UvrABC 核酸内切酶
umuC、*umuD*	编码易错修复所需的蛋白
sulA	编码抑制细胞分裂的蛋白，以争取时间使修复得以进行
recA	编码易错修复和重组修复所需的 RecA 蛋白
与 DNA 代谢有关，但在修复中的作用未知的基因	
SSB	编码 SSB
uvrD	编码 DNA 解螺旋酶Ⅱ
himA	编码位点专一重组、复制和转座所涉及的整合寄主因子亚基
recN	参与重组修复
未知功能基因	
dinB	
dinD	
dinF	

DNA 修复的目的在于保持遗传信息的完整，从而得以存活，但超越创伤复制却易招致差错和常引起突变，这似乎相矛盾。实际不然，这样产生的突变虽最终导致许多细胞的死亡，但少数细胞的 DNA 可能得到修复，有些突变细胞也还能存活。许多细胞的死亡只是为一些细胞的存活所付的代价。

参 考 文 献

1. Blacklurn, E. H. Structure and Function of Tolomeres. Nature. 1991. 350:569～573.
2. Echols, H. and Goodman, M. F. Fidelity Mechanisms in DNA Replication. Annu. Res. Biochem. 1991. 60: 477～511.
3. Fangman, W. L. and Brewes, B. J. A Quesion of Tine: Replication Origins of Eukaryotic Chromosomes. Cell. 1992. 71:363～366.
4. Kornberg, A. and Baker, J. A. DNA Replication. W. H. Freeman and Company, New York: 1992.
5. Lehninger, A. L., Nelson, D. L. and Cox, M. M. Principles of Biochemistry. Worth Publishers, Inc. New York: 1993.
6. Meselson, M. and Stahl, F. W. Equilibrium Sedimentation of Macromolecules in Density Gradients. Proc. Nat Acad. Sci. USA. 1958. 44:671～682.
7. Modrich, P. Mechanicms and Biological Effects of Mismatch Repair. Annu. Rev. Genet. 1991. 25:229～253.
8. Radman, M. and Wagner, R. The High Fidelity of DNA Duplication. Sci. Am. 1988. 259: 40～46.
9. Voet, D. and Voet, J. D. Biochemistry. John Wiley and Sons. New York: 1990.
10. Wang, T. S. F. Eukaryotic DNA Polymerases Annu. Rev. Biochem. 1991. 60:513～552.
11. Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. and Weiner, A. M. Molecular Biology of the Gene. Benzamin/Cummings Publishing Company, California: 1987.
12. Woodward, D. O. and Woodward, V. W. Concepts of Molecular Genetics. Mcgraw-Hill Book Company, New York: 1977.

（张玉麟）

第六章　DNA　重　组

第一节　引　　言

所有的 DNA 都是重组体(recombinant)。DNA 重组(recombination)不仅在减数分裂(meiosis)中发现,而且也在高等生物的体细胞中,衣藻(*Chlamydomonas reinhardii*)的叶绿体基因间和酿酒酵母(*Saccharomyces cerevisiae*)的线粒体基因间发生。在温和噬菌体的整合过程中和转座子(transposon)的转座(transposition)过程中也发生重组。所以,可以说只要有 DNA,就会发生重组。因为生物要生存就得适应,要适应就得有变异(variation),而变异的来源就是突变(mutation)和重组。但就每个个体而言,发生突变的概率毕竟是很低的,涉及到的基因数目也非常有限。所以,如果只有突变而没有不同个体间的基因交换(gene exchange),则生物体难以迅速地组装以产生最能适应环境的基因组合。而且,通过不同个体间的基因交换,可以保证遗传的多样性,从而为选择奠定物质基础。

根据对 DNA 序列和所需蛋白质因子的要求可以把重组分为 4 种类型,即同源重组(homologous recombination),位点专一性重组(site-specific recombination),转座重组(transposition recombination)和异常重组(illegitimate recombination)。同源重组发生在 DNA 的同源序列之间,真核生物非姊妹染色单体的交换,姊妹染色单体的交换,细菌及某些低等真核生物的转化,细菌的转导、接合、噬菌体的重组等都属于这一类型。同源重组要求较大的 DNA 片段进行交换,只要它们的序列相同或接近相同,就可以在序列中的任何一点发生重组,当然也存在重组热点,即某类序列发生重组的概率高于其它序列。大肠杆菌的同源重组需要 RecA 蛋白质。类似的蛋白质也存在于其它细菌中。因此,细菌中的同源重组又称为依赖 RecA 重组。转座子是可移动的 DNA 片段,转座过程中转座单元从染色体的一个区段转移到另一个区段,即发生转座重组,改变了染色体的结构,它既不依靠转座单元和插入区段序列的同源性,也不需要重组酶。转座作用除了单纯地移动基因外,还打乱了 DNA 序列而造成缺失、倒位和重排。这种改变显然是染色体进化的关键特征,特别在真核细胞中更是如此。位点专一性重组是直接在专一序列之间配对而发生的重组,即噬菌体基因组整合到细菌染色体基因组中去,因此,重组只需要有限的同源序列,但必须有位点专一性的蛋白质因子参与催化。异常重组不需要 DNA 序列的同源性,也不需要重组酶参与。目前对这类重组的机制尚不清楚。

可见重组不是偶然的,而是细胞为了编码和调节的必要酶促过程。重组酶类除了提供遗传变异外,当 DNA 被辐射或化学故障损伤时,通过重组可以将同源染色体未损伤 DNA 链置换损伤部分,使细胞损伤的序列恢复过来。而且重组的一些特定类型还调节基因表达。通过染色体上特定片段的互换,细胞能将暂停活动的基因放到能表达的位点上,甚至产生新的蛋白质编码区。

在体外进行 DNA 片段的连接和修饰只是模拟重组和突变的自然过程,遗传学上的重组操作往往是融合和重排染色体。研究重组的分子机制可为基因的熟练操作提供新方法,例如,

利用基因的转座原理，可为细胞的分化和发育提供依据。转座子在染色体上的移动，控制了结构基因的开启，造成了不同细胞中基因启动上的差异，还可启发对癌细胞发生原因的研究。转座子最重要的应用是作为基因定位的标记和通过转座子在不同染色体上的插入和嵌合、转移和克隆基因。通过转座子上的标记基因，就可检测突变基因的位置和分离突变基因。又如同源重组技术是目前能精确地修饰基因组的最有效的方法，使外源 DNA 与受体细胞染色体上的同源序列发生重组，并整合到预定的位点上，就能改变细胞遗传特性。它能根据实验设计对哺乳动物细胞基因组结构进行定量、定点地改变，因而能定向改变细胞或整体本身的遗传结构和特征，这对目前分子生物学及基因工程领域中的主要研究内容：基因结构和功能，基因表达和调控及人类遗传病的基因治疗的研究极为重要。重组技术的发展又为重组机制的研究开辟了广阔的前景。

第二节　同 源 重 组

同源重组也称交换（crossing over），是指减数分裂（meiosis）过程中染色体间遗传物质的交换，即在两个双螺旋 DNA 分子间的相互作用，其特征是重组酶能以两个 DNA 分子中任何一对同源序列作底物进行交换。从整个基因组看，重组频率并不是恒定的，它受综合效应和局部效应的影响。如在精子细胞中同源重组的频率比卵母细胞高出一倍以上，同时重组频率还受染色体结构的影响，例如，在异染色质附近，同源重组就要受到抑制。细菌没有减数分裂，同源重组发生在接合过程中，交配的染色体 DNA 在两个紧密连接的细胞中转移，单细胞内在复制中和复制后的两条同源染色体也可发生交换。同源重组至少具有 3 种明显的功能：一是维持种群的遗传多样性；二是在真核生物中它使染色单体（chromatid）产生瞬间的物理连接，这对第一次减数分裂中染色体正确分离到子代细胞至关重要；三是它有助于 DNA 损伤几种类型的修复。关于前两种功能早已引起研究工作者的极大兴趣，认为同源重组是遗传多样性的来源。而细胞中 DNA 的修复功能也早就引起重视，因为遗传信息贮存于未损伤的互补链中，单链 DNA 的损伤能准确地得到修复。DNA 的损伤有几种类型如双链断裂、双链交联（cross-linking），或在复制中单链的损伤，甚至互补链本身也可能受损等等。当这种情况出现时，准确修复 DNA 的信息必须来源于不相连的同源染色体，因此涉及同源重组。图 6-1 比较了减数分裂期可观察到的染色体的变化和发生物质交换双链 DNA 分子的相互作用。减数分裂开始于较复杂的前期 Ⅰ（prophase），前期可分为 5 个阶段，减数分裂开始的重要标志是染色体变得清晰可见，每条染色体各自复制组成两条姊妹染色单体，每一姊妹染色单体都包括一个双链 DNA 分子，同源染色体相互靠近并在一个或多个区域开始配对，形成二价染色体（bivalant），配对区段的长度取决于同源片段的长短。这一过程称为联会（synapsis）或称染色体配对。联会结束时染色体相互连接形成联合丝复合体（synaptonemal complex）。它们在物种之间存在很大的差异，每种生物都各具其特有的结构。最近从分离酵母突变体中发现其重组步骤与减数分裂很相似。

一、同源重组的机制

（一）断裂-复合

同源重组是指 DNA 分子内断裂-复合（breakage and reunion）的基因交换。对于其作用机制有两种假设：第一种假设认为交换是完整 DNA 分子的断裂重接（breakage and rejoining）。这是从细胞学观察得来的经验图式，即当卷曲的染色体减数分裂和配对时，有时在染色单体的

减数分裂过程	分子相互作用
细线期 凝聚染色体变得清晰	每条染色体已复制，由两条姊妹染色体组成
合线期 染色体开始在有限区或某些区域配对	起始 在一个基因组上发生断裂
粗线期 联会丝复合体沿整个配对染色体伸展	链交换 单链与另一基因组发生链交换
双线期 染色体分离，但在交叉点连接	同化 链交换区域的延伸
终变期 染色体凝聚，从被膜分离；保留交叉，四条染色单体全部变清晰 	拆分 通过切割释放基因组

图 6-1　第一次减数分裂前期的重组

水平上发生物理的断裂。这是染色体收缩所产生的张力造成的。断裂端可通过交叉重接而解除张力，从而产生两个相互重组的染色单体（图 6-2 a），按照这个模式，重组发生在染色体复制完成之后，那就是在每对联会染色体（synapsed chromosomes）中有 4 条染色单体时发生。另一种假设称为拷贝选择（copy choice），即配对的染色体复制时，沿着父本（paternal）染色体形成的新 DNA 链转而以母本（maternal）染色体为模板。假如母本链的互补链当达到相同点时也转换模板，结果就形成两个相对应的重组体（图 6-2 b）。这两种假设的基本区别在于对重组染色体物理起点的推测，前者为两个亲本染色体在断裂和重接之后，产生重组染色体的物理遗传物质，后者则为 DNA 在复制时由于选择不同的模板产生重组。

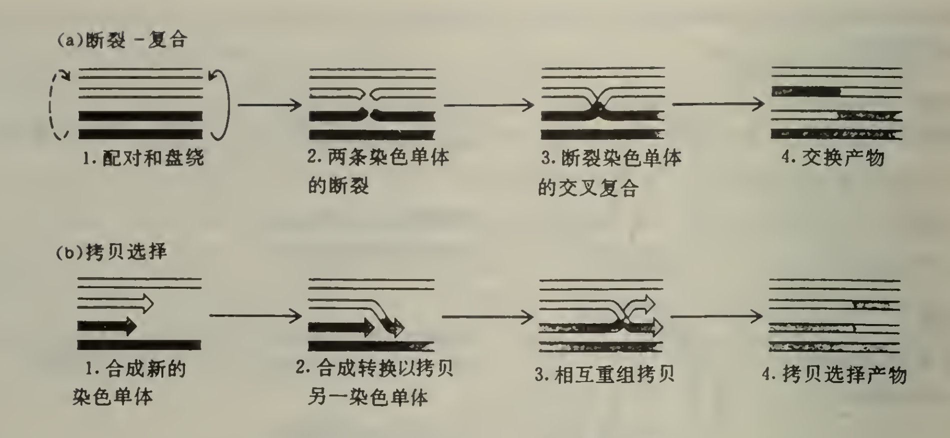


图 6-2 染色体间发生交换的两种假设

交换是 DNA 分子的断裂和重接，它是通过很多实验得到证明的，如图 6-1 所述染色体复制在间期中已经完成，而交换出现在减数分裂的早期，在细菌和高等生物中也可看到同样的现象，例如，在组织培养细胞中常用染色法分辨染色单体，可看到哺乳动物染色单体之间发生明显的 DNA 物理交换。另外，用标记噬菌体进行实验，发现重组与 DNA 复制无关，重组体可以在非复制 DNA 分子之间产生，进一步证实了完整双螺旋的断裂-复合是交换的主要机制。

(二)异源双链

两个 DNA 分子的连接是重组过程的核心，重组从两个配对 DNA 双螺旋同源片段的相应位点上的断裂开始，断裂使由切口产生的自由末端移动，每个单链离开它本身的配偶链到另一个双螺旋上的互补链交叉配对。这种相互交换在两个双链 DNA 分子之间形成连接体。一对相连的双链 DNA 称为连接分子(joint molecule)。一个 DNA 分子上的一股单链与另一个双螺旋的交叉位点称为重组体连接点(recombinant joint)。在重组位点上，每个双链都有一个由两个亲本 DNA 分子中的一股单链共同组成的双链区域，这段区域称为杂种 DNA(hybrid DNA)或异源双链 DNA(heteroduplex DNA)。所以，在两条染色体相连时，中间就有一段异源双链区，遗传分析证明所有病毒都有异源双链区。每个 T 4 分子上异源双链区往往长达几千个核苷酸。λ 噬菌体中的异源双链区的特点也是其长度远比最初配对阶段可达到的长度长。所增加的长度大致都是分支迁移引起的。

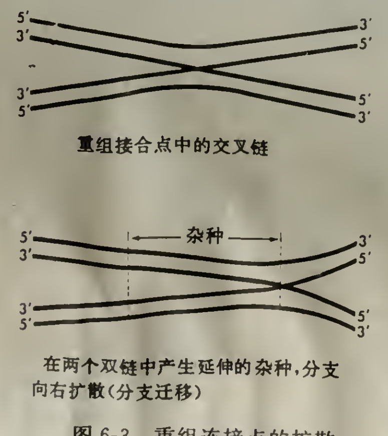


图 6-3 重组连接点的扩散

(三)分支迁移

两个双螺旋形成的交叉连接是两个原来的分子等量碱基交换的所在，交叉连接很容易以拉链式效应扩散，也就是链交换可沿着双链滑动，这个过程称为分支迁移(branch migration)(图 6-3)。特定链的延伸区段可以通过这种方式从一个螺旋移至另一个螺旋而产生异源双链区的杂种 DNA 长片段。分支点可以向任何方向移动。分支迁移具有理论和实践上的重要性。在理论上，它是重组结构的一个重要动力学特征，在实践中，它的存在说明了

体外分子检测不能产生分支点·因为分子解离时分支可发生迁移。分支迁移的速度不易确定，体外观察的结果可能不适应于维持自然条件下大范围异源双链 DNA 的形成，在体内进行大范围的分支迁移必须有重组酶的催化。

（四）Holliday 结构

重组中连接两个 DNA 双链的交换中间物含有 4 股 DNA 链，在其连接处为了转换配对所形成交叉链的连接点(crossed-strand junction)称为 Holliday 结构（图 6-4）。重组分子在 Holliday 连接点(junction)上主链能自由转动，使一条链从一个双螺旋移至另一个双螺旋而不失去碱基对或等价地消除拓扑束缚力。

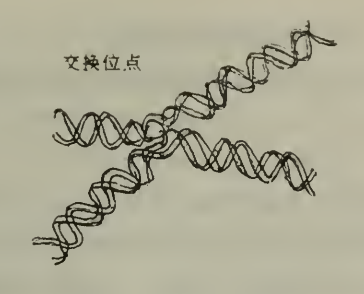


图 6-4　Holliday 结构

双链 DNA 配对

切开同源链

双螺旋间断裂链发生交换

通过分支迁移移动交换点

在相同链上第二次切割

在另外链上形成第二次切口

封住切口

双螺旋间进行第二次链交换，封住切口

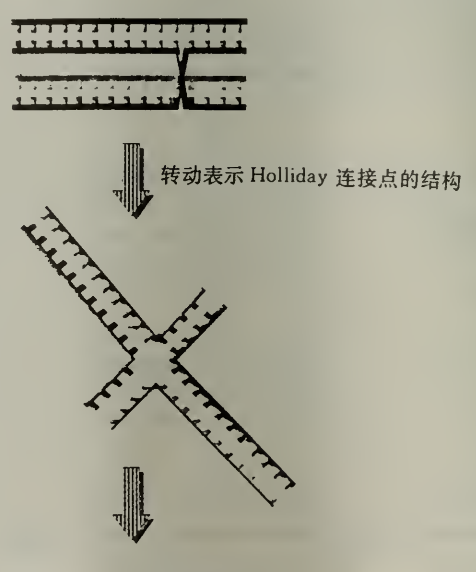
基因组不是重组体，但含有异源双链

产生相互重组基因组

图 6-5　两个配对双链 DNA 之间的重组，从互补交换开始，分支迁移延伸，然后通过切割而拆分

（五）连接分子的拆分

链交换形成的连接分子必须拆分(resolution)，彼此分离为双链DNA 分子。拆分要求两个切口进一步配对，反应结果则取决于配对链上切口的位置。由于交联在一起的两条双链 DNA 分子实际上处在不断地异构化中，切口可能在两对同源链中任意一对上发生(如图 6-5 右)。如果切口发生在形成连接分子过程中原先没有切割过的配对链上，至此所有 4 条单链实际上都已发生过断裂，由此将释放出两条重组 DNA 分子，一条亲本双螺旋分子与另一条亲本双螺旋分子共价相连，中间有一段异源双链区域，这种重组称为交互重组(reci-procal recombination)(图 6-5，右)。如果切口发生在重组过程中曾被切开过的一对同源单链上，那么原来一直保持完整的另一对同源链将继续保持完整。这种切割释放出两条原来的亲本 DNA 分子，并没有发生交互重组，但每个双链 DNA 都保留一段较长的异源双链 DNA(图 6-5，左)。由此可见，连接分子可以以两种交替方式进行拆分，两侧基因或是发生交互重组，或是无重组发生。但不论哪种拆分方式都说明一个原则，即链交换后总会在两条 DNA 分子上留下一段异源双链区，但侧翼序列的重组未必与之同时发生。图 6-6 表示 Holliday 连接点的拆分，连接点上主链自由转动时，根据所形成的一对切口的位置不同，拆分后产生亲本双螺旋或重组双螺旋。



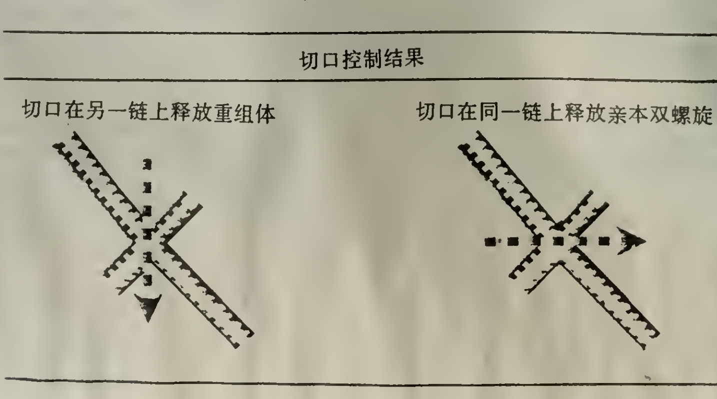


图 6-6　Holliday 连接点的拆分

通过将携带较短的同源序列的质粒或噬菌体导入细菌受体中，就可以观察到重组双链间建立连接体所需最短序列的长度。实验证明，若同源序列小于 75 bp 时，则重组率会明显下降。

二、双链断裂启动重组

图 6-1 的重组一般模式表明，双螺旋 DNA 分子先形成双链切口，从此位点上解链形成的单链可参与并完成基因交换。重组的热点就是双链断裂开始的位点。图 6-7 是目前提出的重组模式。图中把引发重组（双链断裂）的 DNA 分子称为受体，把受体链侵入的同源双螺旋称为供体。当核酸内切酶在受体 DNA 双链分子上切开一个切口时，重组就从此开始。通过核酸内切

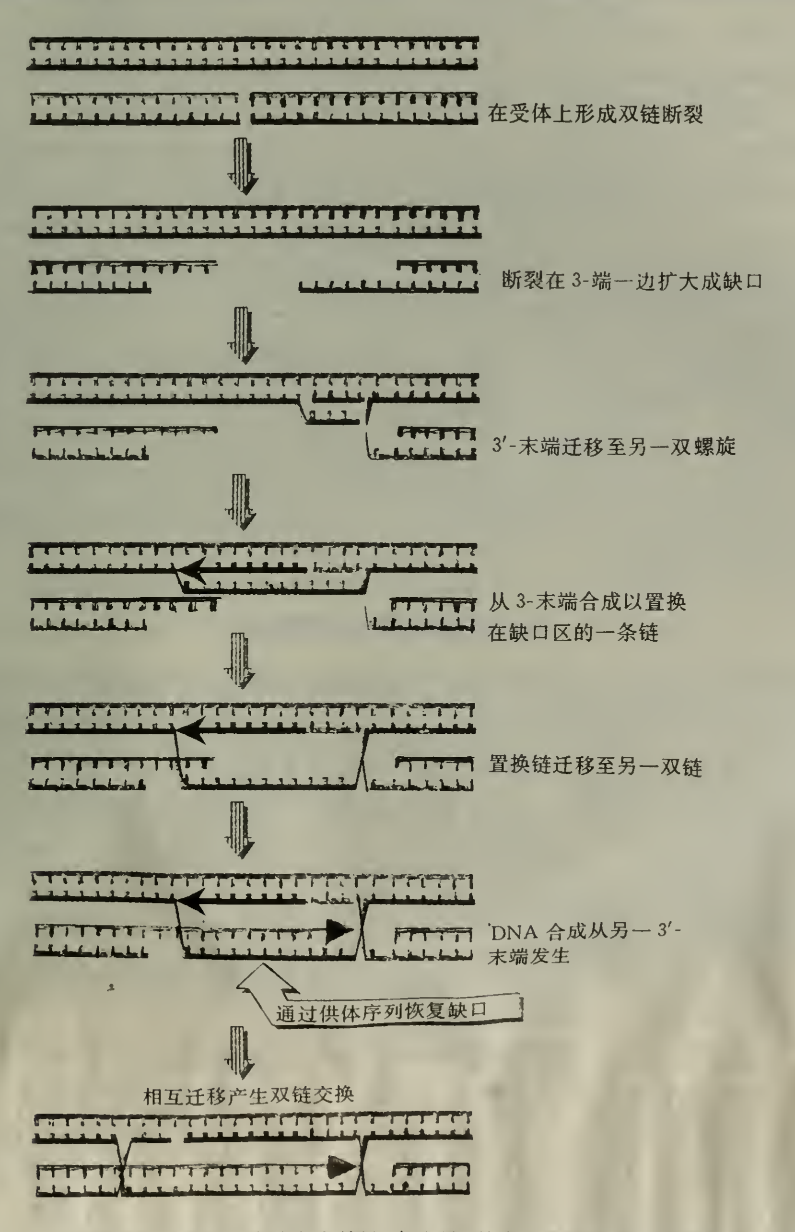


图 6-7　双链断裂启动重组，随后形成单链 3′-末端，其中一条链迁移至同源双链上

酶的进一步切割，扩大成缺口。然后核酸外切酶在断裂的另一端的一条链上切割，产生了3′单链末端。这样，一个游离的3′端侵入供体双链的同源区。异源双链的形成产生了D-环结构，其中供体的一条链被取代。D-环利用3′端作为引物，通过修复合成而延伸，最后，D-环逐渐延长，足以达到交换染色单体上缺口的全长。当突出的单链到达缺口的远侧(尽头)，互补单链序列就退火，这样在缺口的任何一侧都有异源双链，而缺口本身则代表了D-环单链。用缺口左侧的3′-末端为引物，通过修复合成将缺口区修复成完整的双链。总之，缺口是由两股独立的单链DNA的合成途径得到修复的。分支迁移把这个结构转化成具有两个重组连接点的分子，连接点必须通过切割而拆分。如果两个连接点以同样的方式拆分，则释放出原来的两个未交换的分子，它们各自含有一个改变了遗传信息的区段，该区段是足迹试验交换的结果。

如果两个连接点以相反的方式拆分，则是基因交换的结果。拆分前具有两个连接点的分子结构说明了双链断裂模式和只是产生单链交换的模式之间是截然不同的。一是随着双链的断裂，区段各端包括交换区都形成了异源双链DNA，两个异源双链片段的序列相当于缺口区段。此时，在两个DNA分子中缺口区被供体DNA序列所取代(图6-7)。因此，异源双链的序列是不对称的。并且一个DNA分子的部分序列可以转换成另一个DNA分子的对应序列。这就是行使引发作用的染色单体称作受体的原因；二是随着互补单链交换的进行，每个DNA双链从交换起始点到分支迁移点之间的区段都有异源双链成分(图6-5)。单链交换模式不同之处在于一些DNA能降解或重新合成(而起始染色单体是基因信息的供体)。双链断裂模式并未降低异源双链DNA形成的重要性。它仍是双链分子相互作用唯一合理的方式。双链断裂乍看起来令人惊异，因为一旦断裂贯穿整个DNA分子就无法回到原状。比较图6-5和图6-7，在单链交换模式中，没有任何信息丢失，但在双链断裂模式中，在切割开始就有信息丢失，在纠正信息时任何错误都是致命的。另一方面，借助于另一双链进行再合成以恢复失去的信息，可给细胞提供一个主要的安全措施。

三、双链断裂引发联会

多年来一直认为联会丝复合体代表了重组的某个阶段，也可能是DNA交换必不可少的开端。近来则认为联会丝复合体是重组的结果而不是原因，联会开始于每条染色体(姊妹染色单体对)环绕一种称为轴心体(axial element)的结构发生凝聚。轴心体是一种蛋白质，各染色体的轴心体排列在同一直线上。联会丝复合体形成三分体结构，此时联会丝复合体的轴心体称为侧体(lateral element)。每条染色体在此期间呈一团染色质与连着一个侧体的状态(侧体有条纹结构)，两个侧体被一个精细而致密的中体(central element)分开，这些平行而致密的链沿轴盘旋扭结处在一个平面上。从分子水平来看，同源染色体之间的距离较大，超过200 nm (DNA分子直径为2 nm)。所以，尽管同源染色体排列在同一直线上，但是，它们的间距太大，以致不能实现同源DNA分子间的接触。

但是，在真菌或昆虫的球形体或圆柱体结构中，能看到联会丝复合体两侧形成的交叉连接，称为节(node)或重组节(recombination noduce)。它们出现的频率及分布与染色体交叉相同，因此，可证明是重组的位点。关于重组与联会丝复合体形成之间的关系已经确定，在果蝇或酵母中，取消染色体配对的突变可阻止重组的发生。目前没有一个体系将重组的分子机制与细胞活动做出比较。最近，在对酿酒酵母的减数分裂的分析中取得了一些进展，二者的时间关系概括于图6-8。有充分的证据表明在酵母体内无论是同源重组还是位点专一性重组都是双链

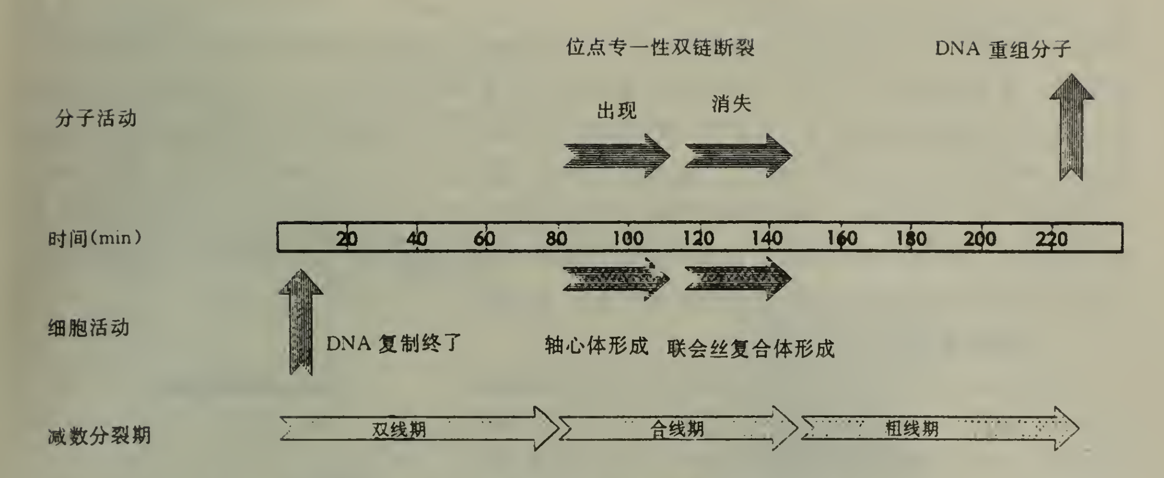


图 6-8　轴心体形成出现双链断裂，而当联会丝复合体延伸时就消失，DNA 重组体可在粗线期末检出，测定时间从减数分裂前的复制终了时计算

断裂启动重组。双链断裂最初在交配型染色体转换过程中发现，在转换过程中一段序列被另一段序列所置换。双链断裂发生在减数分裂早期的特定位点上，这就为重组提供了热点。重组在某一位点附近的频率是高的，在此位点的一侧或两侧，重组频率呈梯度下降，双链断裂所造成的平端两侧同时转变成长的 3′-单链末端(图 6-7)。酵母突变体(rad 50)是重组缺陷型的，它阻断了平端转变为单链突出的粘性末端，这表明双链断裂是重组过程必不可少的。双链断裂出现在轴心体形成时期，消失在配对的染色体转变成联会丝复合体时。这种精巧的时间上的对应表明：联会丝复合体的形成是重组引发的结果，而重组的引发又借助于断裂双链的出现和随后的重组中间体的形成。这一观点由 rad 50 突变体使轴心体无法转变成联会丝复合体的结果得到验证，这与传统的减数分裂的观点形成鲜明的对照，即联会丝复合体是染色体配对以产生重组分子的需要。联会丝复合体的产生与预计的交换时间是一致的。至于重组是否发生在联会期，一直难以提供直接的证据。因为用重组体的出现来确定的重组是在减数分裂结束之后。但是在酵母中，直接根据是用含有限制性位点的 DNA 分子的产生来确定重组体的出现，它证明重组是发生在粗线期的末期。显然，这将重组置于联会丝复合体形成之后。在重组过程中，联会丝复合体可能占据中介的位置，在同源染色体相互识别以后，双链断裂产生单链末端，并参与寻求互补序列，寻求的结果使同源片段之间连接，这种连接使它们紧密接近并引发三叉联会丝复合体的形成。在双线期染色体脱去了联会丝复合体，于是交叉就成为染色体上可见的连接点，表示基因交换的发生。但是高叉分子的性质尚不甚清楚，它可能代表交换完成的残余物，或代表同源染色体之间的连接点，而基因交换尚未拆分。在减数分裂末期，交叉移向染色体的两端，这种变化说明它代表重组过程的某种残迹，并非真正的重组中间体。

四、细菌 DNA 的重组和单链同化

要研究 DNA 分子间序列交换的本质，必须藉助于细菌体系。虽然识别反应只是重组机制的一小部分，而且是局部的，只涉及到 DNA 分子有限的区域，并非整个染色体，但是分子相互作用的一般顺序都是相似的：从分子断裂产生的单链与其亲本双螺旋作用；配对区域不断扩大；核酸内切酶拆分亲本双螺旋；很多酶参与重组的各个阶段(尽管这些酶只代表部分重组

酶)。细菌重组酶是通过在其基因中存在的 *rec*⁻ 突变而发现的,*rec*⁻ 突变体表现型丧失了重组能力而不能进行一般的重组。细菌不进行双链大量交换,在原核生物体内有几种不同的途径引发重组。在某些情况下 DNA 可有效利用单链的自由 3′-末端参加反应(为接合作用)。单链的缺口可以受外部影响而产生,当 DNA 受到紫外线辐射、X 射线辐射及化学药剂等作用时双链断裂或产生缺口(或双链的单链区),这些处理都会使重组增加。而对细菌本身而言,当其DNA 连接酶和 DNA 聚合酶 I 的基因发生突变,复制和修复中的 DNA 就不能正常封闭,从而产生断裂和缺口。单链的尾部则由正在进行滚动复制的噬菌体基因组产生。如果有两个双链 DNA 分子参与交换,则必然要产生单链区和自由的 3′-末端。

(一)细菌重组酶

在重组反应中其活性最为明显的是由大肠杆菌染色体上 *rec* 基因和 *ruv* 基因编码的一些酶。RecA 蛋白具有联会同源 DNA 分子的能力,并启动一个分子的单链侵入到另一分子的双螺旋上。通过取代双螺旋上的原始链而形成异源双链,RecBCD 核酸酶结合在 DNA 分子 chi 序列的一端,沿着该链移动时解开 DNA 双链,并在 *chi* 位点上切割而产生单链末端。RuvA 和 RuvB 指导异源双链 DNA 的形成,而 RuvC 则具有切开 Holliday 结构的功能。

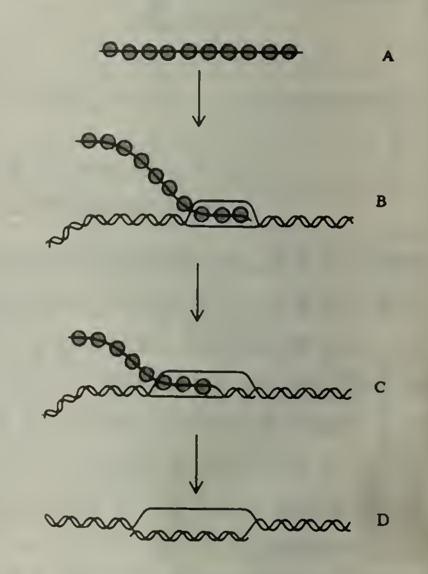


图 6-9 用体外 DNA 片段的模式反应表示 RecA 蛋白的活性

A. RecA 蛋白首先结合在单股 DNA 上;B. 双链 DNA 的不同区段被结合到单链片段上的 RecA 蛋白解链;C. 一旦发现同源性就开始退火。RecA 蛋白延长了部分退火的杂交链;D. 完成杂交,形成 D 环。

1. RecA 蛋白　对重组机制了解得最为详细的是大肠杆菌 *recA* 基因产物的研究。靶染色体的单链 DNA 与其互补序列的匹配需要 RecA 蛋白的催化,RecA 蛋白是催化重组基本反应的酶,能通过碱基配对使两个 DNA 连接成杂交分子。RecA 蛋白引发单链 DNA 与其互补双链之间的配对,能用单链去置换双链中的同源链(图 6-9),RecA 蛋白能在体外催化 Holliday 结构的形成,图 6-11 说明 RecA 蛋白如何在体外从两个相同的双螺旋(其中之一在末端有缺口)形成完整的重组体。当 RecA 蛋白驱使退火到缺口边缘时(图 6-10 c)缺口上部螺旋就解旋,这样连接点扩散到整个双螺旋区,同时交换点的右边形成第二个杂交双螺旋(图 6-10 d)。RecA 蛋白在这一阶段还表现出第二种作用,即它指导形成四股螺旋。四股螺旋由交换点左边区域的两个 DNA 双链组成,从而加速了链交换,使连接通过整个双螺旋而移动。图 6-11 A 的结构与图 6-10 d 的完全双链部分相同,但有两点不同:一是两个同源双链到交换点的左边画成四股螺旋;二是RecA 蛋白存在于四股螺旋区。图 6-11 B 表示 RecA 蛋白从四股螺旋中间体引发互补的链交换。

2. RecBCD 酶　亦称核酸外切酶 V,它通过 *chi* 序列识别靶位点,这种酶的各个亚基是由 *recB*、*recC* 和 *recD* 基因编码的产物。它是一种有效的降解 DNA 的核酸酶,具有多种活性,可以在单链结合蛋白的存在下松开 DNA 双链,同时还具有 ATP 酶活性。它在重组中的作用是提供具有 3′-末端的单链区域。遗传学实验证明,核酸外切酶 V 以相当高的频率促进含有 *chi*

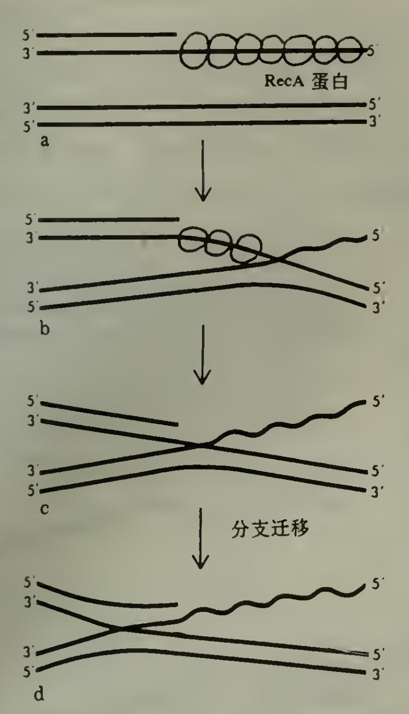




图 6-10 RecA 蛋白在体外催化形成重组体的模式反应

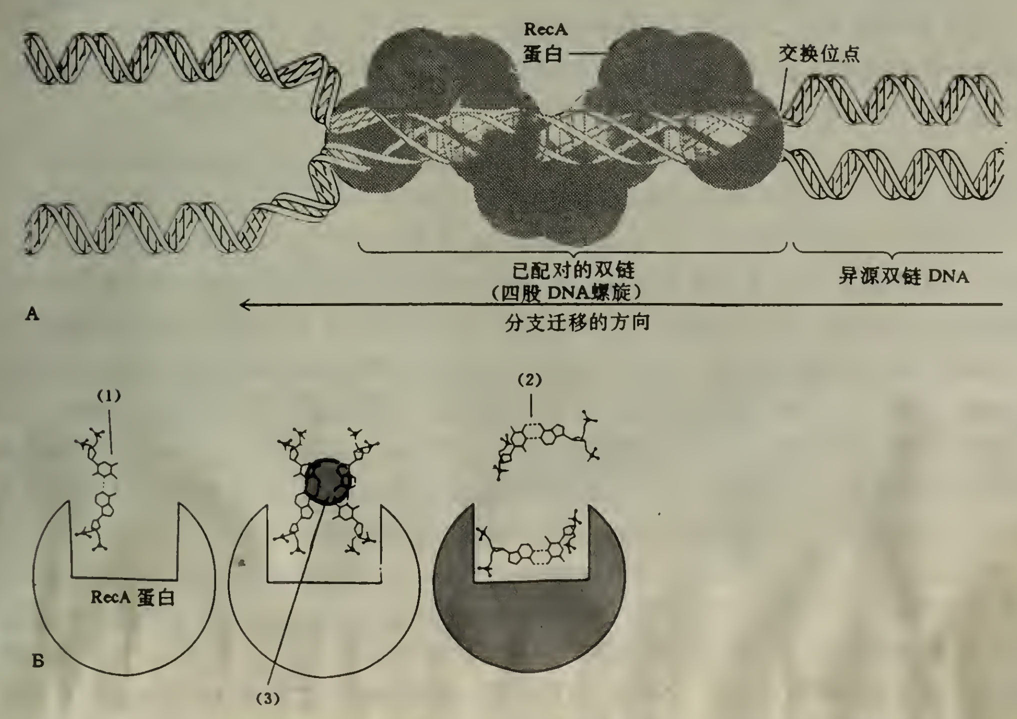


图 6-11 A. 重组中的 DNA 四股螺旋；B. (1)双螺旋结合在 RecA 蛋白上；(2)链转换以形成两个异源双链；(3)两条 DNA 双链间形成四股螺旋假设的配对位点

位点的 DNA 重组，*chi* 位点具有恒定的 8 bp 不对称序列 $\begin{matrix}5'GCTGGTGG3'\\3'CGACCACC5'\end{matrix}$ 在自然情况下这段序列在 *E. coli* 中每隔 5～10 bp 出现一次。*chi* 序列可以在大范围内激发重组，有的甚至远在 10 kb 处起作用。当核酸外切酶Ⅴ结合于 DNA 的 *chi* 位点右侧时，便可沿着正在解链的 DNA 移动。当它移动到 *chi* 位点右侧 4～6 kb 时，在上边的链上将单链切断。核酸外切酶Ⅴ的作用见图 6-12。这个模型是体外反应所得数据建立起来的。它参与解旋、切割和产生引发形成异源双链连接点的末端。

3. 在重组后期发挥功能的一组酶　在大肠杆菌中存有由 3 个基因 *ruvA*，*ruvB* 和 *ruvC* 组成的一组酶，其中 *ruvA* 和 *ruvB* 两个基因产物会促进异源双链结构的形成。这种结构是在 RecA 蛋白催化一个双螺旋和一个单链反应所形成的。RuvA识别 Holliday 结构连接点，RuvB 是一种 ATP 酶，为分支迁移提供动力，分支迁移是以 10～20 bp/s 速度进行。第三个基因是 *ruvC*，是编码专一识别 Holliday 结构连接点的核酸内切酶，它在体外切断连接点以拆分重组中间体。

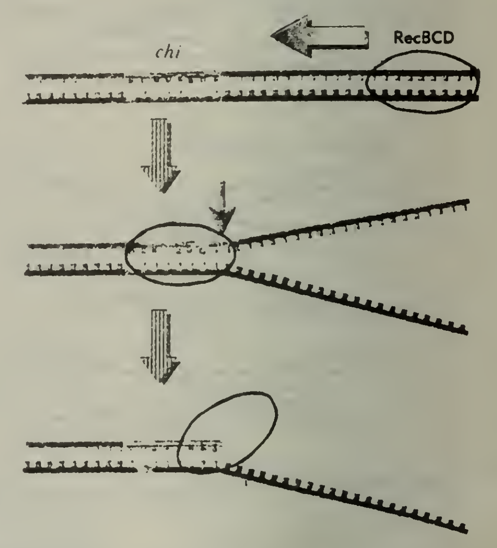


图 6-12　RecBCD 核酸酶从右侧到达 *chi* 序列，并在 *chi* 侧切断

（二）单链摄入和链交换

RecA 蛋白作用于 DNA 分子，用单链去置换双链分子的同源区段，这个反应称为单链摄入（single-strand uptake）或单链同化（single-strand assimilation）。置换反应可以在几种不同构型的 DNA 分子之间发生，但必须具备下列 3 个条件：①其中一个 DNA 分子必须具有单链区域；②其中一个 DNA 分子必须具有自由 3′-末端；③单链和 3′-末端必须位于 DNA 分子间的互补区域内。反应以两种形式进行，当一线型单链侵入环型双链时，在互补区内它取代了原先的一条链并形成 D 环。当一环状单链去作用于线型双链时，链交换产生了开环分子。图 6-13 中，反应的 DNA 分子是一个环状单链和部分与环状单链同源的线状双链片段，双链片段的一条链被环状单链的同源片段所置换。单链同化与重组的引发有关，各种重组模式都要求有中间体形成，在中间体上，一条或两条单链从一条双链上交换到另一条双链上（图 6-5 和图 6-7）RecA蛋白催化中间体的形成。RecA 蛋白能将单链和双链 DNA 聚集为长的细丝，表明它具有促进分支迁移的能力，在每个细丝的转弯处结合 6 个 RecA 蛋白单体，每个单体结合三个核苷酸，细丝具有螺旋结构。

DNA 分子的相互作用就发生在这些细丝中，当一条单链被双螺旋同化时，首先由 RecA 蛋白将单链结合进入细丝，然后双螺旋也被结合进来形成三链结构。在这个体系中，联会先于 DNA 的物理交换，因为配对反应可以在缺少自由末端的情况下发生，而此时的链交换却是不可能的。链交换需要自由的 3′-末端，反应在细丝中进行，RecA 蛋白始终结合在最先形成的单链上，因此，在反应结束时，RecA 蛋白就结合到双链分子上。在反应进行过程中，大量的 ATP 会被水解。ATP 的作用是让 RecA 蛋白对 DNA 更具有很大的亲和力，这是结合和配对反应的需要。ATP 水解后就会使结合部位的亲和力下降，有利于异源双链 DNA 的释放。可将 RecA

蛋白催化单链和双链 DNA 的反应划分为 3 个阶段:①缓慢的联会前阶段,此时 RecA 蛋白结合在单链上;②单链与双螺旋的互补链迅速配对,形成异源双链连接分子;③从双螺旋结构中缓慢置换一条链产生一段长的异源双链 DNA。

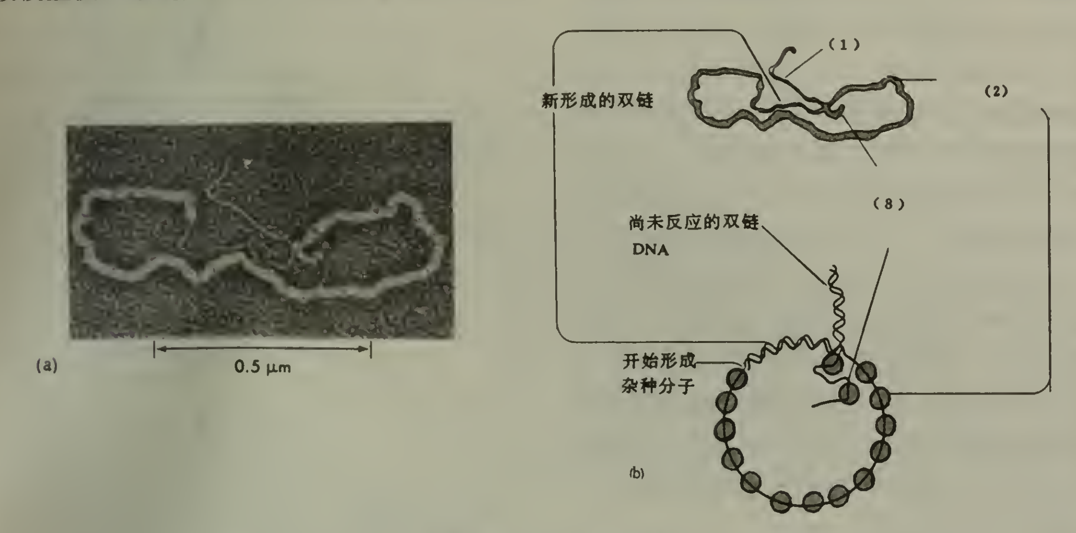


图 6-13 RecA 蛋白催化体外链交换的电镜图(a)及示意图(b)

(1)尚未反应的双链 DNA(同电镜图);(2)单股环状 DNA 与 RecA 蛋白复合;(3)从双链置换下来的 RecA 蛋白结合 DNA 链(因为有过剩的 RecA 蛋白可利用)。

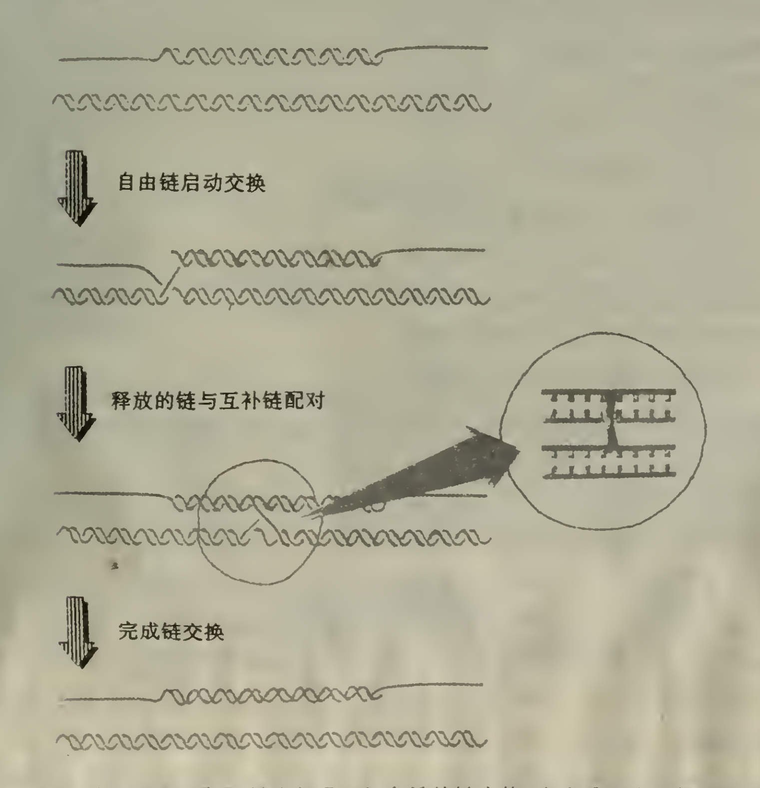


图 6-14 在部分双链或全部双链之间 RecA 介导的链交换,产生重组中间体连接分子

单链侵入还需要单链结合蛋白(SSB)，以保证底物不形成二级结构，从而促进反应的进行。SSB 与 RecA 蛋白都结合在伸展的 DNA 上，表明它们在单链同化中的作用是结合 DNA 以便形成一个与细丝相连的结构。当单链分子与双螺旋作用时，双螺旋分子在重组连接区段开始解旋，异源双链开始的区域不是以传统的双螺旋形式存在，而是两条链并排地连到一起，这种类型的区段称为平行汇接(paranemic joint)，平行汇接是不稳定的，随着反应的进一步深入，它就转变成双螺旋形式。这种反应是以等摩尔消除负超螺旋方式进行的。这就需要一种酶通过瞬间断裂让一条链环绕另一条链转动来完成松旋和紧旋。以上讨论的即单链侵入到双螺旋中，只是重组反应的一部分。

在链交换过程中(图 6-14)，一部分双螺旋分子和一个完整的双螺旋分子作用，单链同化开始于分子的末端，入侵的单链以常规的方式取代了双螺旋上的同源片段。当反应进行到两分子都是双螺旋时，入侵的单链从其互补链上分离下来，互补链再与其它被置换的单链配对。在这个阶段中 DNA 分子形成重组连接体，在体外 RecA 蛋白催化 Holliday 结构的形成，表明它介导了互补链的转换。但是，要完成链交换还需要另外 3 个基因产物参加，RuvA 识别 Holliday 结构的连接点。RuvB 是一种 ATP 酶，可以为分支迁移提供动力，加速分支迁移向前推进。RuvC(识别 Holliday 结构的核酸内切酶)切割 Holliday 结构的连接点，拆分重组中间体完成链的交换。根据各种蛋白质的催化作用，可以说明大肠杆菌重组过程的各个阶段(图 6-15)：利用重组从另一条双链取得 DNA 修复双链的缺口。细菌中的反应大多发生于 DNA 片段与整条染色体相互作用中，并且 DNA 损伤促使了修复反应的发生，它们完全不同于减数分裂期间基因组之间的重组，但在控制 DNA 方面则具有相同的分子活性。RecA 的同系物在原核生物中普遍存在，在真核生物中也发现了相关蛋白质。例如，酿酒酵母中 DMC 1 和 RAD 51 两个基因编码了与 RecA 相类似的蛋白质。这些基因突变会引起一种类似的表现型，即双链片段不断积累，即无法形成正常的联会丝复合体。说明 DNA 双螺旋之间的链交换与联会丝复合体的形成有关，并且增加了染色体联会与细菌单链同化反应有关的可能性。但是，真核生物中 RecA 同系物并不形成细丝，因此其反应可能有所不同。

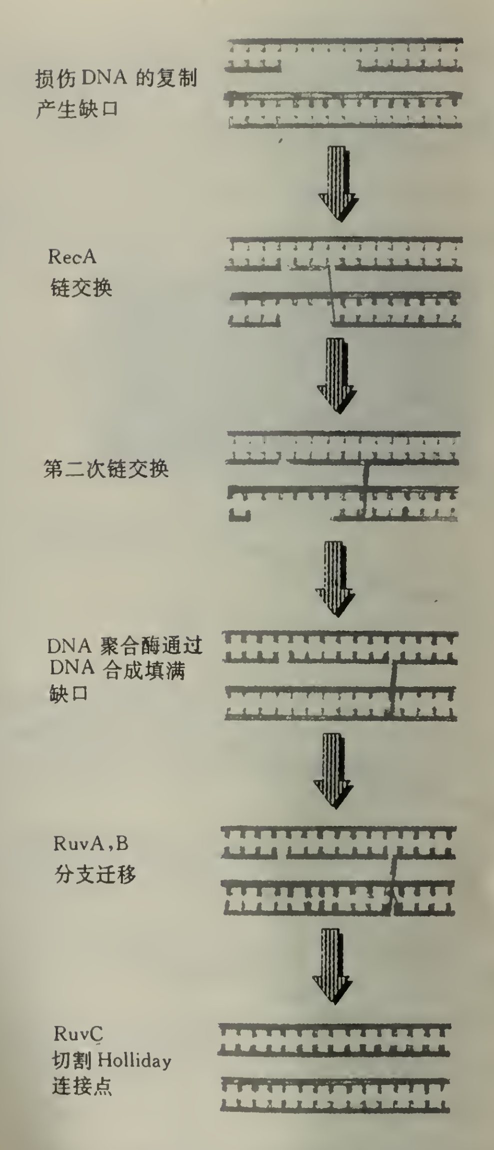


图 6-15 修复途径中随着适宜底物 DNA 分子的产生，细菌酶能催化重组的所有阶段

五、基因转换导致等位基因间重组

异源双链 DNA 参与重组是等位基因间发生重组的特征，等位基因重组也推动了异源双链模型的发展。几乎有 1/4 的单基因形成异源双链，其本身就是异源双链的原因，发生在基因座内的个别断裂和复合产生了相互配对的重组染色体。因为在子囊菌(*Ascomycetes*)中，每一次减数分裂的产物包含在一个大的子囊里，所以，用子囊菌做材料来研究重组的机制较为方便。用减数分裂产生的四倍体细胞核呈线性排列，减数分裂发生于这 4 个细胞核产生之后，并形成 8 个线性排列的单倍体细胞核。图 6-17 中这 8 个单倍体细胞核分别代表减数分裂产生的 4 个染色体上 8 条链的遗传特征。杂合子经减数分裂后每个等位基因产生 4 个拷贝。在大多数孢子中都可见到这种情况。但也有一定比例的异常孢子，在等位基因不同区域，异源双链的形成和校正是孢子比例异常的原因。图 6-16 说明重组过程中，4 条经减数分裂的染色体中的一条存有一段长的杂种 DNA，可能是由于双链断裂而引发了重组。假设两个等位基因由于单点突变而产生差别，当发生链交换产生异源双链时，异源双链的两条链会在突变位点上产生错配。实际上，DNA 的每条链都带有不同的遗传信息，如果序列上未发生变化，那么，两条链就会在保真复制中发生分离。每条链产生了信息不变的双链，这种现象称为减数分裂后分离(post-meiotic segregation)，因为它反映了减数分裂后的链分离。它的重要性在于直接证实了等位基因重组过程中异源双链的存在。在等位基因重组时也发生了另一种情况：等位基因的比例不同于原来的 4∶4 的比例，这种现象称为基因转换(gene conversion)。它说明一条染色单体与另一条染色单体信息间的非相互重组。DNA 分子的链交换导致了基因转换，而序列的改变又在分子水平上提供了基因转换的原因之一。其过程为：①图 6-7 所示的双链断裂模型，一个 DNA

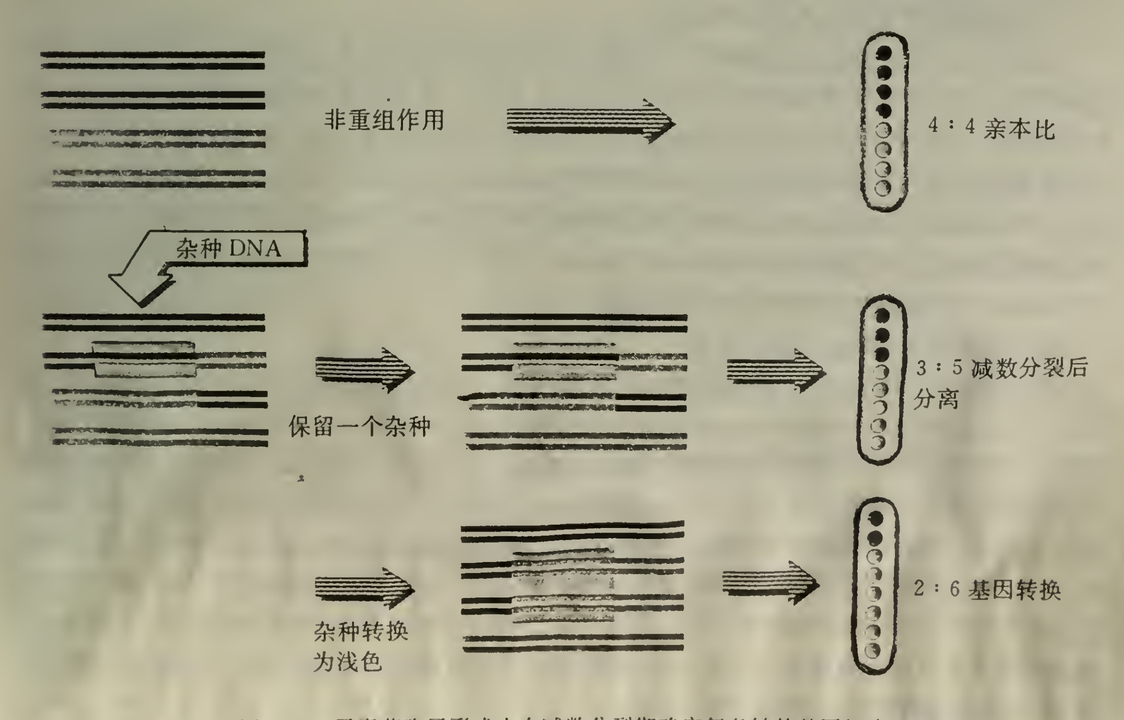


图 6-16　子囊菌孢子形成中在减数分裂期确定每条链的基因组成

分子作为遗传信息的供体通过链交换反应和缺口填补反应取代了互补双链上的对应序列；②异源双链 DNA 作为交换过程的一部分，在一个双螺旋上的单链与其另一螺旋互补区域配对时产生的。当修复系统识别异源双链上错配的碱基对时，切除和置换其中一条链使其恢复互补性。这种过程使代表一个等位基因的 DNA 链变成了代表另一个等位基因的序列。

基因转换并不取决于交换，但与交换有关。大部分异常的子囊在等位基因转换位点两侧的标记之间出现遗传重组。因为如果异常比例是由于重组引起的结果(图 6-5 和图 6-7)，那么无论重组与否，结构的拆分频率大致相等(图 6-6)，意味着真菌染色体所引发的交换频率大约相当于远距离基因重组频率的两倍。目前有各种不同观点企图从分子水平来解释重组。一种观点认为基因转换的两个方向或者效应相同，或者等位基因向一个方向的转换优先于另一方向的专一性效应；另一种观点则认为重组频率梯度从热点(双链断裂开始的位点)两侧递降。重组频率梯度形成与热点处缺口的扩大并转化为较长的单链末端有直接关系。基因转换的某些信息，得自许多基因簇成员的序列研究，重组产物往往会发生分离从而不适用于 DNA 序列水平上的分析，但是，如果某基因簇两个成员之间发生不等量交换，这两个非等位基因之间就可形成异源双链。基因转换有效地将其中一个非等位基因转换成另一个基因序列。在同一条染色体上如果存在一个以上的基因拷贝，便可用足迹法(foot printing)以追踪其活动。如异源双链的形成和基因转换超过一个基因还多一部分，那么这部分序列可以与另一基因完全相同或紧密相关。而剩余部分则呈现较大的趋异性。有效的序列在基因转换过程中延伸很远，甚至达到数千个碱基对。

六、重 组 修 复

(一)交换中产生的插入和缺失差错

大多数的交换过程是非常精确的，它依赖于重新形成氢键时互补单链区的正确匹配。这种正确排列为大多数多核苷酸序列的独特性所保证，因为一条多核苷酸的随机链若多于 10～12 个核苷酸时，它几乎不会与相似长度的其它片段具有相同的序列。因此，只要单链片段比较长，一个基因的片段与其本身错误部分或与不同基因的错误连接是极不可能的。但是在极少数的情况下，当两条不同染色体的区段具有很大同源性时，发生偶然的错连也是可能的。结果导致一大段核苷酸的插入或缺失(insertion or deletion)(图 6-17)。这种偶然的重排往往由 RecA 蛋白所催化，可能还有其它未知的酶在起作用。

(二)重组修复损伤的 DNA 分子

我们常认为交换是产生遗传多样性的机制，但它主要的功能还在于修复(repair)DNA 的损伤。这是在研究 recA 细菌和酵母重组缺陷型突变体(recombination-defective mutant)中新发现的。因为它们对射线和一些化学药剂非常敏感致使 DNA 受损。DNA 受损后所产生的缺口通过重组修复形成重组中间体，单链缺口被同源重组的 DNA 填满后就完成修复。但修复不仅是否切去中间体以交换两个螺旋的侧臂，重要的是以正确的序列填补缺口。虽然简单缺口可由 DNA 聚合酶来填满，但如缺口含有胸腺嘧啶二聚体，它不能进行碱基配对，复制时必须在受伤位点前停止，在后面再重新开始。伤痕上的遗传信息已从两条 DNA 链上丢失，只有通过同源双链重组才能得到恢复。尽管一个细菌细胞往往有其染色体的其它拷贝，但新复制的姊妹 DNA 分子是正确序列最容易得到的来源。

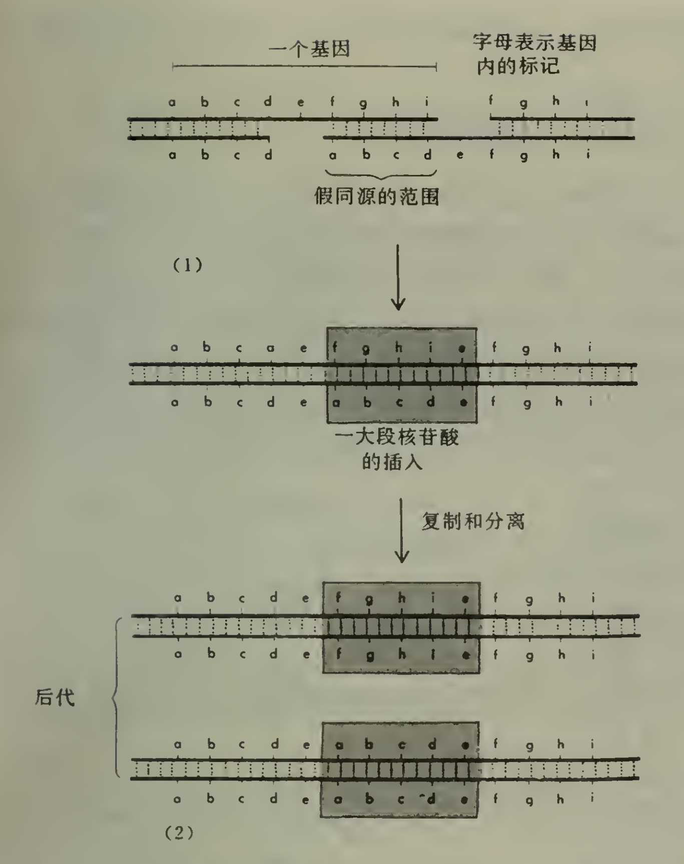


图 6-17 相同基因内近似同源序列的配对导致多核苷酸的插入
(1)相同的基因在不同部分中的近似同源区段间的配对形成了异源双链;(2)两个后代分子各含有不同的一大段插入的核苷酸。

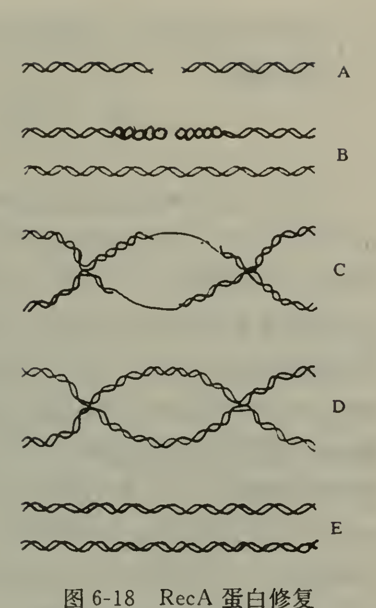


图 6-18 RecA 蛋白修复断裂双链 DNA 的模型
A. DNA 中丢失了一个片段,造成双链断裂;B. 解链之后,被核酸酶降解,在各断裂端暴露出单链 DNA,RecA 蛋白与之结合,并启动与同源 DNA 链交换;C. 形成一个双重组连接点;D. DNA 聚合酶填补产生的缺口;E. 两个重组连接点均已拆分,产生两条完整双链。

上述机制的引伸可以说明 RecA 蛋白如何修复最基本的损伤,如由于 X 射线照射引起的双链断裂。核酸酶使每个断裂形成单链末端,单链末端能分别入侵一同源序列以形成两个邻近的重组连接点(图 6-18)。此外,RecBC 酶进入每一个自由末端以提供起始位点。当两个结合点切开时,分离了亲代螺旋,所产生的缺口由 DNA 聚合酶填补,并由 DNA 连接酶封闭。

第三节 位点专一性重组

位点专一性重组需要在特定位点上发生断裂和重接,从而产生精确的 DNA 重排。与同源重组不同,位点专一性重组不是由 DNA 序列的同源性(虽然也涉及短的同源序列)所导引,而是在结合序列部位由专一的酶来催化断裂重接。位点专一性的重组是由能使相应酶出现的调节过程所引发,而同源重组则是由能与 RecA 蛋白结合的 DNA 序列随机断裂引发的。

位点专一性重组最早是在 λ 噬菌体的遗传学研究中发现的。λDNA 通过重组整合到大肠杆菌染色体的专一性位点上,转变成为原噬菌体 DNA 的非活性状态。一般位点专一性重组都具有 λ 整合作用的两个基本特征:①交换是相互的和保存原先的 DNA,是典型的保守性重组(conservative recombination);②它发生在噬菌体和细菌 DNA 短同源序列中的专一性核苷酸

上，在高等生物细胞中，最重要的例子是通过一组前体序列的位点专一性重排构建多种多样的专一抗体基因。

一、位点专一性重组的机制

λDNA 在不同生活型之间的转换包括两种状态。在溶菌状态时，λDNA 在宿主细菌中以环状分子独立存在，而在溶源状态时，噬菌体 DNA 则是细菌染色体的整合部分，称原噬菌体。这两种状态的转换都与位点专一性重组有关。要进入溶源状态，游离的 λDNA 必须整合(integrate)到宿主 DNA 中去，为了从溶源状态过渡到溶菌状态，原噬菌体 DNA 必须从细菌染色体上切除(excise)下来。整合和切除过程是在细菌 DNA 和噬菌体 DNA 专一的附着位点(attachment site 或 *att*)上进行的。细菌的附着位点称 *attB*，由序列 BOB′ 组成，噬菌体的附着位点称 *attP*，由序列 POP′ 组成。图 6-19 说明在这些位点处发生的重组反应。其中 O 序列是 *attB* 和 *attP* 所共有的，所以，称为核心序列(core sequence)。重组就发生在这段序列上。侧翼区 B，B′ 以及 P，P′ 称为臂(arm)，其序列则各不相同。由于噬菌体 DNA 是环状的，在重组过程中，它以线性形式插入细菌染色体中，此时原噬菌体被重组产生的两个新 *att* 位点所固定。所以，形成这两个新的附着位点 *attL*(BOP′)和 *attR*(POB′)是很重要的，这样可以使整合和切除反应不在一对相同的反应序列之间发生。整合过程要求在 *attB* 和 *attP* 之间进行识别，而切除过程则在 *attL* 和 *attR* 之间识别。因此，位点专一性重组反应的定向特征取决于重组位点的识别。尽管重组反应是可逆的，但反应导向则由不同条件决定，这是噬菌体生活过程的重要特征，因为它保证了整合反应不会立即被切除反应所逆转。反之亦然。

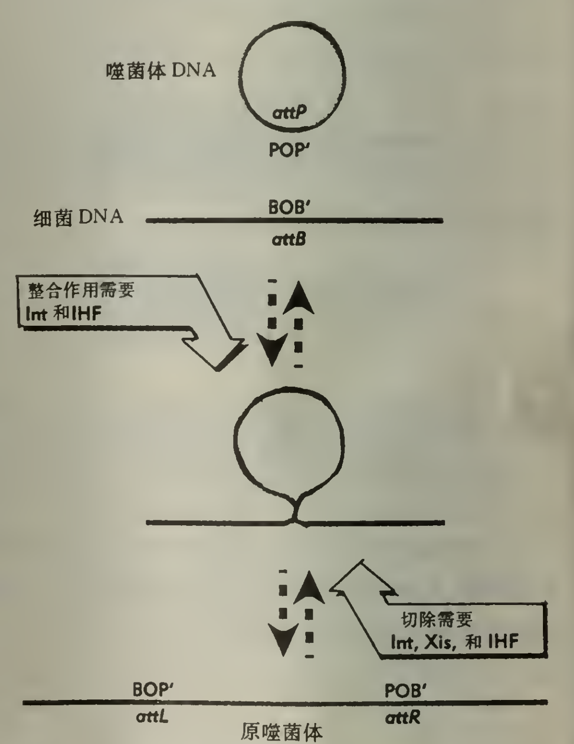


图 6-19　通过在 *attP* 和 *attB* 间的相互重组，环状的噬菌体 DNA 转换为整合的原噬菌体，原噬菌体通过 *attL* 和 *attR* 间相互重组而切除

(一)整合酶

λ 噬菌体编码一种酶——λ 整合酶(λ integrase)，后者直接将噬菌体 DNA 插入大肠杆菌染色体中，通过两条 DNA 的专一位点上的重组合并成一个环状分子(图 6-20)。在噬菌体侵染早期产生较多的整合酶，所以，整合作用几乎发生在每个被侵染的细胞中。整合反应还需要由大肠杆菌编码的整合宿主因子 IHF(integration host factor)的共同作用，IHF 是由两个不同亚基组成的蛋白质，分子量为 20 kD，分别由 *himA* 和 *himD* 基因编码。*him* 基因的突变会阻止 λ 位点专一性重组的发生，同时 λ*int* 基因突变也会抑制 *him* 基因突变的发生，这表明 IHF 和 Int 的相互作用。

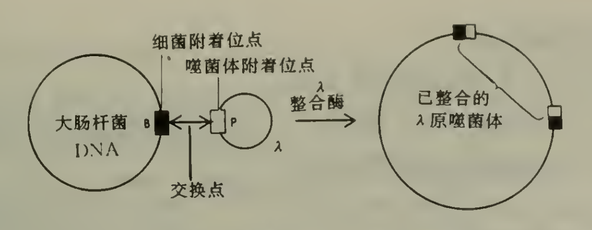


图 6-20　通过位点专一性重组，λ 噬菌体 DNA 整合到大肠杆菌染色体中

（二）整合反应

离体研究链交换的进程可利用自杀底物(suicide substrate)的方法去中止中间体的形成。此时在核心序列上造成切口，干扰了重组的进行，有利于识别重组反应已经开始但未结束的 DNA 分子。中间体结构表明了链交换发生的顺序。Int 蛋白能切断 DNA，并使它重新连接，进而使 Holliday 结构拆分。图 6-21 模型说明 *attP* 和 *attB* 部位进行的交叉断裂，互补单链末端进行交叉杂交。λ DNA 交换点之间的距离为 7 bp，反应产生出 3′-磷酸末端和 5′-羟基末端，反应说明了所产生的重叠单链末端进行了退火。每个螺旋的相应链在同一位置上切割，游离的 3′-末端在双螺旋之间进行交换，7 bp 的序列沿同源区分支迁移，然后在对应的双链上切割并拆分。离体反应需要在 *attP* 上有超螺旋，而 *attB* 则不需要，Int 蛋白具有松弛负超螺旋的能力。体外重组反应需要大量的 Int 和 IHF 蛋白，它们共同结合在 *attP* 上，由于超螺旋加强了它们对结合部位的亲和力，这种高剂量的蛋白并不引起酶促作用，而是为了维持单链重组所要求的结构。

噬菌体和细菌 DNA 组合

噬菌体 DNA

P-GCTTT TTTATACTAA-P′
CGAAAAAATATGATT

B-GCTTT TTTATACTAA-B′
CGAAAAAATATGATT

细菌 DNA

交错切割导致交叉配对

封闭重组体连接点，产生整合的原噬菌体 DNA

B-GCTTT TTTATACTAA-P′
CGAAAAAATATGATT

P-GCTTT TTTATACTAA-B′
CGAAAAAATATGATT

图 6-21　在 *attP* 和 *attB* 共同的核心序列上交错切割，通过交叉复合产生相互复组体连接点

（三）*att* 位点的结构分析

整合酶的结合位点，是通过测定其真实交换位点需要多少个毗连的 DNA 碱基，才能满足整合酶的结合来确定的。定界试验说明 *attP* 远远大于 *attB*，*attP* 需要 240 bp 才能行使其功能，而 *attB* 只需要从 $-11\sim+11$ 的 23 bp 的片段。其中在核心区两侧只有 4 bp，*attP* 和 *attB* 的 bp 数不同，暗示两者在重组中起的作用不同。与位点专一性重组有关的蛋白质结合在 *att* 部位的特定位点上，Int 蛋白有两种不同的结合模式，一种是结合在核心序列的反向位点上，这些位点具有共有序列（consensus sequence），以便在每条链的适当位置上进行切割（图 6-22）；另一种是结合在 *att* 臂上，这些位点没有共有序列。Int 蛋白的不同结构域识别不同类型的序列，含 N-末端的结构域识别 att 臂，含 C-末端的结构域则识别 *attP* 和 *attB* 核心序列。这两个结构可能同时结合于 DNA，促使臂位点靠近核心区。

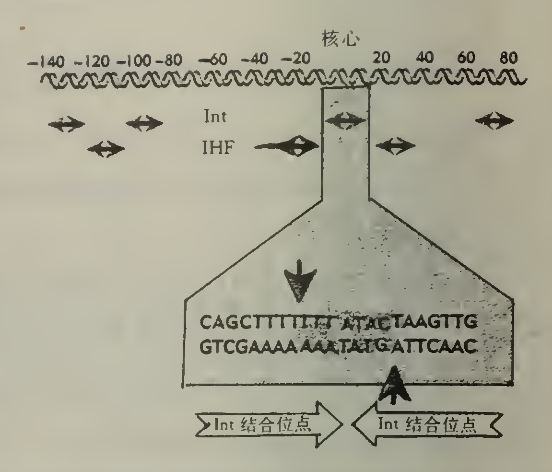


图 6-22　Int 和 IHF 结合在 *attP* 不同位点上，Int 在核心区的识别序列包括切割位点

（四）整合体及其作用

当 Int 和 IHF 蛋白结合到 *attP* 位点时形成的复合物称为整合体（intasome）。在整合体中所有结合部位可以共同引向一寡聚蛋白的表面，形成类似核小体的结构，所以整合体的形成需要 *attP* 位点的超螺旋。*attB* 的结合部位是核心区的两个 Int 位点，但 Int 并不是在游离的 DNA 形成时直接结合在 *attB* 上，整合体就是捕获 *attB* 的中间体。Int 蛋白分子可能作为整合体的一部分结合到 *attB* 的核心位点上，说明 *attP* 和 *attB* 的最初识别不是直接取决于 DNA 的同源性，而是由 Int 蛋白识别两个 *att* 序列的能力来决定的，然后两个 *att* 位点被结合到整合体结构预定的定向部位上，当需要链交换反应时，同源序列在此阶段才显得重要。

（五）切除反应

当诱发原噬菌体生长时，整合作用（integration）发生逆转，这个过程称为切除。Xis 蛋白在控制反应方向上起重要作用。切除需要 Xis 蛋白。它抑制整合作用，Xis 蛋白结合在 *attP* 区相互靠近的位点上，其保护区延伸到 30～40 bp，Int、Xis 和 IHF 连在一起覆盖了全部 *attP*，Xis 蛋白的连接改变了 DNA 的组织结构，从而成为整合反应的惰性底物。整合和切除的不对称性，可通过加入 Xis 之后 Int 与 *attR* 形成类似复合物来说明，这个复合物与 Int 在 *attL* 形成凝聚复合物相匹配。可见反向重组反应必须有 Int 和 Xis 复合体紧密与 *att* 位点紧密结合才能实现。

位点专一性重组的高度复杂性可能是由于调节反应的需求引起的，所以，当噬菌体进入溶源状态时优先发生整合反应。而当原噬菌体进入裂解周期时则切除反应占优势，通过控制 Int 和 Xis 的存量发生适当的反应。

二、位点专一性重组调节基因表达

在一个 DNA 分子上两个位点之间的重组会得到两种结果，这取决于作用位点的取向。重组可以除去插入片段或使片段倒转（图 6-23）。细胞有时利用重组的倒转以在两个交替排列的

DNA 之间做出选择，从而使两种或两组不同的蛋白质得到表达。这种机制常常调节生物外表的蛋白质。例如，噬菌体 *Mu*（增变基因）的尾部蛋白是由倒转的 *gin* 片段调节的，还有沙门氏杆菌（*Salmonella*）的鞭毛抗原亦与此类似。

1. 锥虫（*trypanosome*）表面蛋白的改变则是通过 DNA 连续重排达到的 重组或许用来控制表面蛋白，因为可以使转换完全以使有机体只制造出一种表面抗原，这样就避免了抗原对抗体的轻微反应

2. 沙门氏杆菌的相变（phase variation）导致两种鞭毛蛋白 *H*1 和 *H*2 的交替表达 在特定时间内细胞只表达其中的一种，两种蛋白质不能同时表达。*H*2 基因的启动子位于相邻的、长 970 bp 可倒转的 DNA 片段上，片段的边界为两个方向相反的重复序列，它们的长度各为 14 bp（图 6-24）。短同源序列可以像 λ 整合时那样作为同源序列的核心指导重组。因为同源重复序列的指向相反，因此，片段在每次交换中就倒转。在一种情况中，启动子位于 *H*2 的一边，*H*2 就和相邻的、编码鞭毛蛋白 H1 阻遏物的基因同时转录，结果 *H*2 得到转录而 *H*1 被阻遏。如果片段倒转，*H*2 因没有启动子而不再形成，同时 *H*1 的阻遏物也不能形成，从而产生 H1 蛋白。倒转序列本身编码催化倒转反应的 Hin 酶，当细胞生长受到妨碍时，Hin 酶的活性或表达或增加，因而就形成另一种新的表面蛋白。

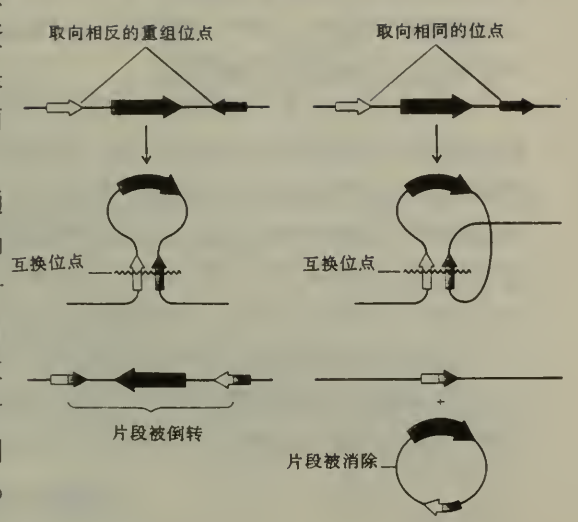


图 6-23 在一个 DNA 分子上两个位点之间的重组，使插入序列倒转或缺失

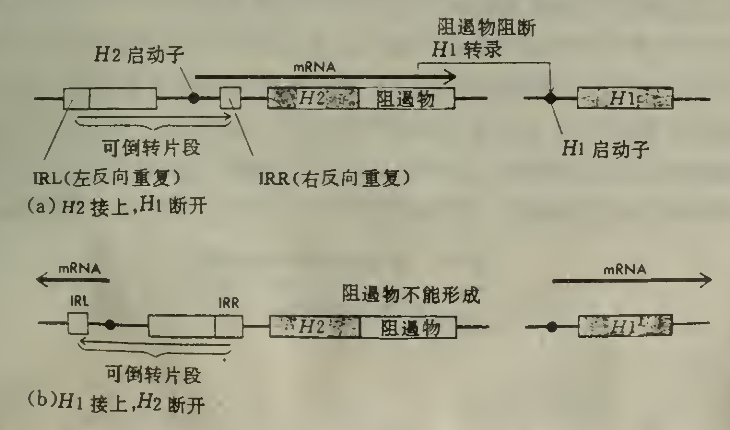


图 6-24 倒转的 DNA 片段控制沙门氏菌鞭毛基因 *H*1 和 *H*2 的交替表达

第四节 转 座 重 组

基因组可通过获得新序列而进化，也可通过现存序列的重排实现这一目的。新序列可通过一定的载体导入基因组。细菌的质粒通过接合作用转移遗传信息，而噬菌体是靠侵染来传递信息；质粒和噬菌体都可偶然地随着它们自身的复制子而转移进寄主的基因组；DNA 的直接转

移可以借助转化作用在细菌间发生。在真核生物中，一些病毒(逆转录病毒)可以在一个侵染循环中转移遗传信息。另外，重排也可以产生新的序列，如果通过转座作用把现存序列放到新调节位置而改变现存序列的功能。

无论原核生物还是真核生物，转座子或转座元件(transposon or transposable element)的移动提供了基因组变化的潜在可能。转座子就是基因组上不必借助于同源序列就可移动的DNA片段，它们可以直接从基因组内的一个位点移到另一个位点。和大部分其它的途径不同，转座作用与供体和受体部位之间的序列没有任何关系。转座子有的是编码蛋白质的序列，这些蛋白质能直接操纵DNA；有的转座子则是和逆转录病毒有关，逆转录病毒的移动是靠它们的RNA转录产物产生的DNA拷贝，然后，DNA拷贝整合到基因组的新位点上。

在原核生物和真核生物中都发现经由DNA移动的转座子。每个细菌转座子携带有编码其自身转座作用所需的酶的基因，可能还需要细菌细胞的一些辅助因子(如DNA聚合酶和旋转酶)的协助。真核生物中也存在类似的系统，目前对参与转座作用的酶的功能还不甚清楚。

一、转座子的类型和结构特征

在原核生物中已发现两种类型的转座子，简单转座子(simple transposon)和复合转座子(composite transposon)，前者称为插入序列(insertion sequence，IS)。所有转座子都有两个结构特征：①两端有20～40 bp的反向重复序列(inverted repetitive sequence，IR)；②具有编码转座酶(transposase)的基因，这种酶催化转座子插入新的位置。复合转座子除了转座过程必需的基因外还含有一到几个基因，它们常有可鉴定的遗传标记如抗菌素抗性或毒素生成等。

(一)插入序列

这类转座子最初是在细菌操纵子中以自发插入形式鉴定出来的。插入序列可以在不同的复制子间转移位置。以非正常重组的方式从一个位点插入到另一个位点，从而对该位点的基因的结构与表达产生多种遗传效应，也常对插入位点后面的基因功能表达产生极性效应。插入序列是最简单的转座子，每一类型都以字头IS加鉴定类型的编号来表示，IS元件是细菌染色体和质粒的正常组分，是一个自主的单位，每种IS元件均编码自身转座所需的蛋白质。每种IS元件具有不同的序列，但有共同的组织形式。大部分转座子插入到宿主DNA的不同位点，但一些转座子不同程度地优先地插入到特殊的热点上。常见的IS元件的参数见表6-1。

表6-1　某些插入序列的参数

元　件	长度(bp)	末端反向重复(bp)	靶位点反向重复(bp)	靶位点选择
IS1	768	23	9	随意
IS2	1 327	41	5	热点
IS4	1 428	18	11 或 12	$AAAN_{20}TTT$
IS5	1 195	16	4	热点
IS10R	1 329	22	9	NGCTNAGCN
IS50R	1 531	9	9	热点
IS903	1 057	18	9	未知

图6-25是插入序列IS 1的结构，IS 1含有转座酶基因，负责IS 1的移动，该基因的两端为24 bp的反向重复序列。

(二)复合转座子

IS 1本身没有任何表型效应。而复合转座子除了与它的转座作用(transposition)有关的基

因外，还含有其它基因，如带有抗药基因，这就赋予宿主细菌以一定的表型，使宿主细菌对某些药物呈现抗性。在抗药性质粒上最早发现带有抗药性基因的转座子，它是一种较大的转座子，又称易位子(translocon)，是由两个重复序列夹着携带抗药物标记的中央区段(含一个或多个结构基因)组成的，通常以 Tn 和后面加数码来命名。有的转座子的重复顺序就是IS。复合转座子也和 IS 一样，能从一个位点转座到另一个位点，或从一个复制子转座到另一个复制子。复合转座子的特性总结于表 6-2 中。

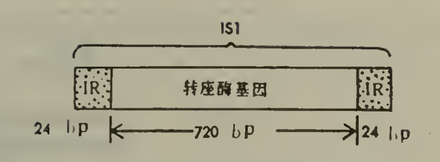


图 6-25　插入序列 IS 1 的结构

表 6-2　几种复合转座子的特征

元　件	长度(bp)	遗传标记	末端组件	组件取向	组件间关系	组件功能
Tn903	3 100	kan^R	IS903	反向	完全相同	两者都有功能
Tn9	2 500	can^R	IS1	正向	可能相同	可能有功能
Tn10	9 300	tet^R	IS10R/IS101	反向	2.5%差异	有功能/无功能
Tn5	5 700	kan^R	IS50R/IS501	反向	1 bp 变化	有功能/无功能

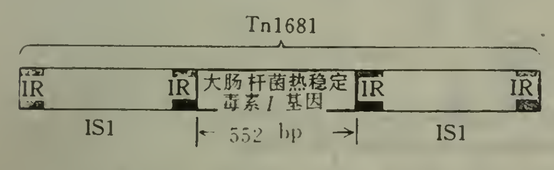


图 6-26　复合转座子结构示意图

在某些情况下，复合转座子的序列组件(module)是完全相同的，例如，Tn903(IS903 的反向序列)和 Tn9(IS1 的正向序列)。在另一些情况下，复合转座子的组件是密切相关的，但不完全相同，如 Tn10 和 Tn5。一个功能性的 IS 组件可以自身转座，也可使整个复合转座子进行转座。当一个复合转座子的组件完全相同时，如可在 Tn903 或 Tn9 中，任何一个组件都可以引起转座子的移动。当组件不同时，如在 Tn10 或 Tn5 中，转座必须严格地完全依赖于其中一个组件。

复合转座子都比 IS 大，约 4 500 bp～20 000 bp。通过将带有 Tn 的 DNA 先行变性，再进行链内复性，在电镜下就可以看到一个有双链的茎和单链的环的 DNA 分子。研究证明单环是 Tn 的结构基因部分，双链茎是两端的一对重复序列。这些重复序列多数是反向重复，因此，在复性时形成双链的茎。用这个方法已测知一些转座子中的结构基因及其两端重复序列的长度。复合转座子的结构如图 6-26 所示。图中两个 IS1 元件构成的转座子 Tn1 681，两个 IS1 组件之间是另一个基因(大肠杆菌热稳定毒素 *I* 基因)。

二、转座子的转座机制

转座子都具有编码与转座作用有关的酶——转座酶的基因，而末端大多数都是反向重复序列。转座酶既识别转座子的两末端，也能与靶位点序列结合。转座作用的机制是转座子插到新的位点上(图 6-27)产生交错切口，所形成的突出单链末端与转座子两端的反向重复序列相连，然后由 DNA 聚合酶填补缺口，DNA 连接酶封闭切口，交错末端的产生与填补说明了靶 DNA 在插入位点存在正向重复，两条链上切口之间的交错取决于正向重复的长度，因此，每个转座子所特有的靶重复序列，反映了切割靶 DNA 的酶的几何形状。

在常见各种转座方式中，尽管交错末端的利用是共同的，但转座作用的机制仍可分为 3 种不同类型，即复制型、非复制型和保守型。

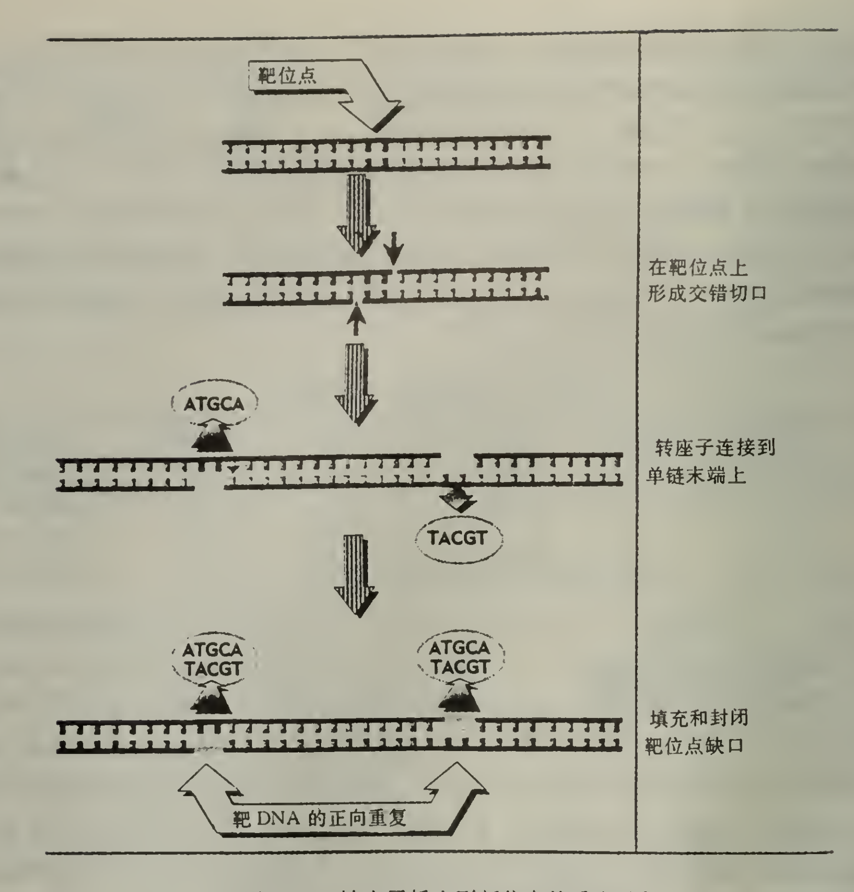


图 6-27　转座子插入到新位点的反应过程

(一)复制型转座

在复制型转座(replicative transposition)中,发生转座的实体是原先元件的拷贝,即转座子作为可移动的元件被复制,一个拷贝保留在供体的原来部位不变;另一个拷贝则插入到受体的位点上,供体和受体都有一个转座子的拷贝,所以,转座过程中扩增了转座子的拷贝数。复制型转座需要两种酶:转座酶作用于原来转座子的两个末端,解离酶(resolvase)作用于复制后的拷贝上。

(二)非复制型转座

在非复制型转座(nonreplicative transposition)中,转座子作为一个物理实体直接从供体的一个位点移动到受体的新位点处,这就会在供体位点上留下一个裂口,其结果是使供体基因组受到轻微的损伤(在具有多拷贝的细菌中可以忍受)或致死,还有一种可能,就是宿主修复系统能识别双链断裂处,并加以修复。非复制型转座只需要转座酶。

(三)保守型转座

保守型转座(conservative transposition)是另一种非复制型转座。转座过程中,转座子从供体位点上切除,通过一系列反应插入到靶位点上,其中每个核苷酸键都保留下来。这种保守型转座反应过程正好与 λ 整合机制相似,而且这种转座子的转座酶与 λ 整合酶家族有关。利用这种机制不但可以介导转座子自身的转移,而且也可以介导供体 DNA 从一个细菌转移到另一个细菌中。所以,把保守型转座的元件称作附加体(episome)更为合适。

实际上,有些转座子仅使用一条途径进行转座;而有的则可能使用两种途径。IS1 元件和

IS903 元件就是使用非复制型和复制型途径。Mu 噬菌体借助于两种并通过一个普遍存在的中间体进行转座。

三、转座子的转座模式

(一)复制型转座模式

复制型转座经由两个阶段。首先,含有转座子的复制子与没有转座子的复制子,通过复制子融合产生一种称为共合体(cointegrate)的结构。共合体含有转座子的两个拷贝,每个转座子都位于原来两个复制子之间的连接处,并且取向相同。然后,转座子两个拷贝之间的同源重组可以产生原先供体复制子,并释放出靶复制子。靶复制子获得的转座子两侧带有宿主靶序列的短重复序列。这一反应称为拆分,由解离酶介导的拆分反应释放出两个各自带有一个转座子拷

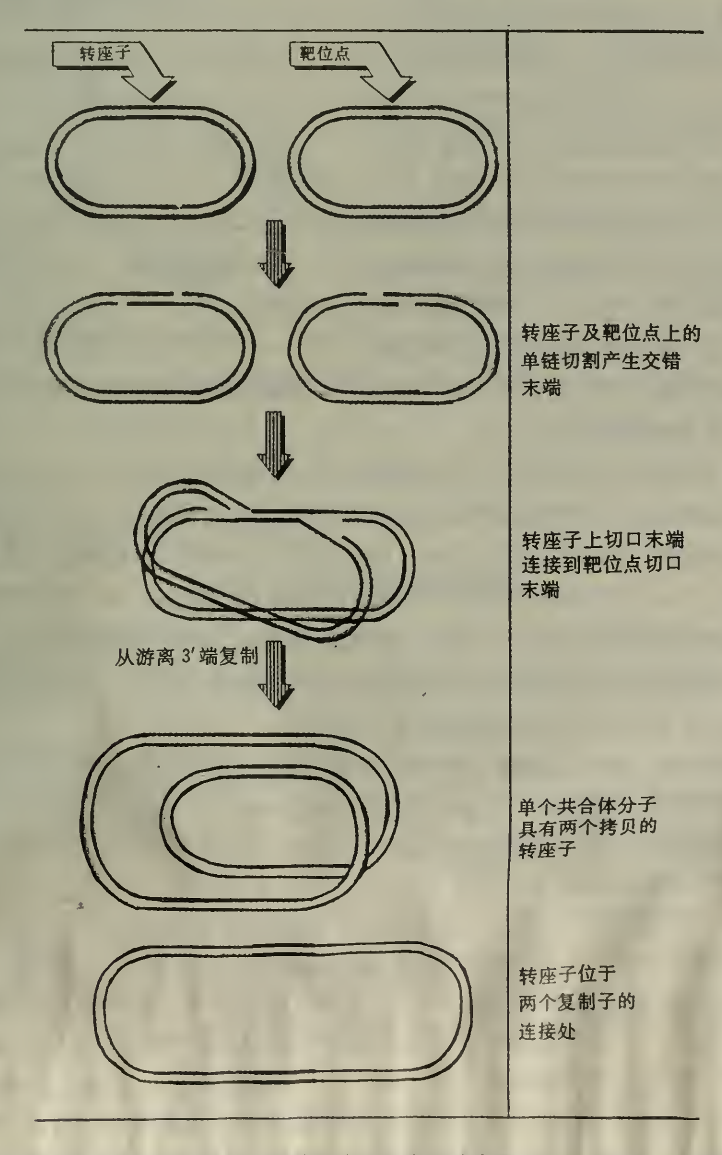


图 6-28　复制型转座形成的共合体

贝的独立复制子。图 6-28 说明 Mu 转座作用中共合体结构的形成。这一过程是由 4 条单链断裂开始的。供体分子的转座子被转座酶切开，该酶具有识别末端位点的专一性，并在靶 DNA 分子上产生 5 bp 的交错切口。供体链在切口处与靶链连接，即转座子序列的每一末端都与在靶位点产生的一条突出单链相连，在连接点上形成一个交叉形结构。交叉形结构的形成是由转座酶完成的。这种结构决定了转座方式。所以，复制型转座是通过在供体位点和靶位点上产生转座子的拷贝，其产物是共合体。

（二）非复制型转座模式

交叉结构也可用于非复制型转座，这种机制的基本原理是通过断裂和复合反应来重新构建靶位点，而供体仍保持断裂，因而不能形成共合体。图 6-29 说明 Mu 噬菌体发生的非复制型转座的切割反应。非复制型转座是通过切割释放出一个交叉结构时产生的。由交错切割所产生的单链区通过修复系统填补。该反应形成的产物是靶复制子，转座子已被插入到原先单链切割处所产生的重复序列之间。供体复制子上产生双链断裂，断裂处就是转座子最初的位置。总之，只要在靶 DNA 上形成切口，在转座子任何一端发生双链断裂，由于双链断裂阻止复制的完成，就迫使反应以非复制型转座的形式进行下去。

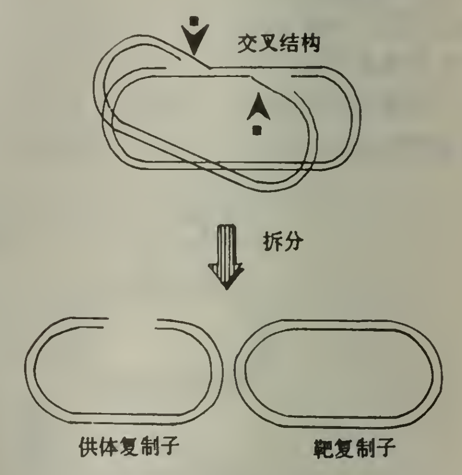


图 6-29　非复制型转座模式

（三）TnA 及其转座模式

TnA 转座子家族是复合转座子，含有较长的核苷酸片段（约 5 000 bp），是复制型转座。它们携带负责自身转座的转座酶基因和诸如抗药性基因等其它基因。TnA 家族转座子的两端没有 IS 或类 IS 的组件，而具有长度为 38 bp 的反向重复末端特征；在靶位点处产生一个 5 bp 的正向重复序列。TnA 介导的转座作用的两个阶段分别是在由 *tnpA* 基因编码的转座酶和由 *tnpR* 基因编码的解离酶的作用下完成的，解离酶需要一个特异的内部位点，这是 TnA 家族独有的特征。转座酶识别位于转座元件末端 38 bp 反向重复中的 25 bp 序列，并在靶 DNA 5 bp 上进行交错切割产生切口，这样以便转座子的插入。

解离酶具有双重功能：这种蛋白既能作为基因表达的阻遏物，又提供解离功能。*tnpR* 基因突变增加了转座频率。原因在于 TnpR 蛋白阻遏了 *tnpA* 及其自身基因（*tnpR*）的转录。因此，TnpR 的失活就增加了 TnpA 蛋白的合成，从而导致了转座频率的增加，这说明转座酶的含量是转座作用的限制因子。

tnpA 和 *tnpR* 基因的表达是一个富含 A—T 的顺反子间（intercistronic）控制区段向外散发的，这可以从图 6-30 的 Tn3 图谱中看出。其两端具有反向末端重复，一个内部 *res* 位点及 3 个已知基因。转座过程必须经过共合体这一中间体；在复制型转座过程中，共合体原则上能被转座子两个拷贝中的任何一个相应的配对点之间的同源重组而拆分，这种共合体结构中的正向重复序列间（即转座子）的重组需要解离酶 TnpR 的参与。Tn3 的拆分反应仅发生在特定位点上。这个拆分位点被称为 *res*（resolvation site），位于 *tnpA* 和 *tnpR* 的中间。如果 *res* 缺失，会引起共合体的积累；在无 *res* 情况下，拆分反应可以为 RecA 蛋白介导的一般重组作用代替，但效率很低。

TnpR 解离酶结合的部位已由 DNA 蛋白复合物足迹法确定，其位置见图 6-30 下部。结合发生在 3 个独立的部位，每个部位长 30～40 bp。3 个结合部位具有序列的同源性，这种同源性说明了它们都有一个双重对称的共有序列。部位 Ⅰ 包括遗传上定义为 *res* 位点的区段。它的缺失，将使拆分反应无法进行；此外，拆分反应也包括结合部位 Ⅱ 和 Ⅲ，因为部位 Ⅱ 和 Ⅲ 这些位点的任何一个缺失，拆分反应都变得很弱。部位 Ⅰ 与 *tnpA* 转录起始位点相重叠，部位 Ⅱ 与 *tnpR* 转录起始位点相重叠。解离酶需要在 3 个部位上结合，才能保持 DNA 合适的拓扑学结构，这样才能在不引起任何 DNA 变化的情况下抑制 *tnpA* 和 *tnpR* 的转录。

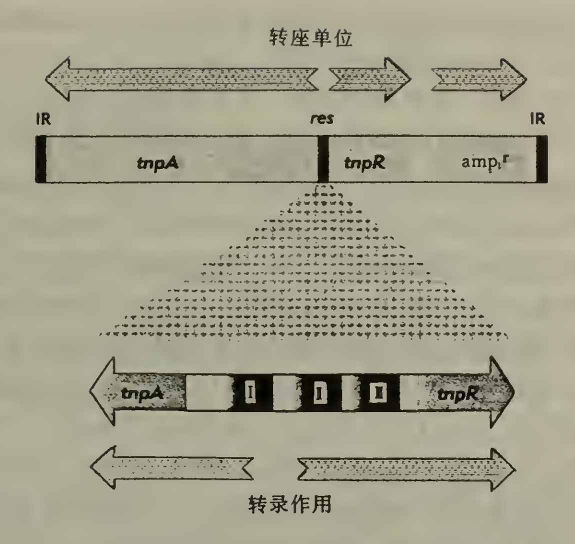


图 6-30　Tn3 转座子的结构

解离酶催化拆分反应是在一个较大的核糖核蛋白结构中发生，解离酶以 4 个亚基结合到重组 *res* 位点上行使功能。每个亚基产生一个单链断裂；然后彼此相关的亚基又重新组织排列，移动 DNA 链，使它们处于一种重组的构象中，接着封闭切口。在此过程中，键的断裂和连接不需要输入能量，解离酶共价附着在 *res* 位点形成的双链切口的 5′-末端上。断裂对称地出现在回文对称区段，产生两个碱基的单链延伸。部位 Ⅰ 交换区域的切割反应为：

$$\begin{matrix}5'\text{TTATAA}3' \\ 3'\text{AATATT}5'\end{matrix} \longrightarrow \begin{matrix}5'\text{TTAT} \\ 3'\text{AA-蛋白}\end{matrix} + \begin{matrix}\text{蛋白-AA}3' \\ \text{TATT}5'\end{matrix}$$

该反应与 λ 噬菌体在 *att* 位点的反应很相似。

四、转座子转座频率的调控

细胞中转座频率的控制是很重要的，为了存活，转座子必须以某一最低的频率移动；频率太高，也会对细胞造成损害，细胞内的转座酶在一般情况下浓度很低，极少引起转座，每个细胞每个世代产生的转座酶分子还不到一个（如 Tn10 转座酶是以每个世代每个细胞中合成 0.15 个分子的低水平进行的），引起自发转座频率只有 10^{-7}次。如果由含有 Tn10 转座酶基因的工程质粒，就能提供更多的转座酶，Tn10 的转座频率可以提高 1 000 倍以上，每个转座子似乎都有控制其自身转座频率的机制，目前对 Tn10 的各种调控机理已研究得很清楚。

Tn10 是一个复合转座子，其中只有末端的几个碱基是转座反应必需的。Tn10 是优先选择特定靶序列（热点）的几个转座元件中的一个。由转座产生的 DNA 两侧的 9 bp 的正向重复中，有一个 6 bp 的共有序列对称地排列在靶位点上，因此，Tn10 两侧重复序列经常采用 $\begin{matrix}5'\text{NGCTAGCN}3' \\ 3'\text{NCGATCGN}5'\end{matrix}$ 的形式，N 代表任何碱基，热点越强，共有序列也就越接近，这很可能是同种转座酶识别靶序列及转座子末端所要求的。

IS10R 是 Tn10 的活性组件，IS10L 组件是功能缺陷型的，它仅能提供 IS10R 转座酶活性的 1%～10%，尽管 IS10L 是缺陷型的，但它的末端能被转座酶所识别（IS10R 和 IS10L 分别

是构成复合转座子 Tn10 两端的两个组件，又称右臂和左臂)。图 6-31 概括了 IS10R 的结构，靠近 IS10R 边界外缘有两个方向相反的启动子：弱启动子 P_{IN}负责 IS10R 的转录，转录方向是向左进行，其转录产物是一个具有编码区段的 RNA，其长度一直延伸到整个 IS10R，并产生 47kD 的蛋白质(即转座酶)；强启动子 P_{OUT}转录向右进行，转录产物称为 OUTRNA，转录可到达邻近宿主DNA 一侧，通常终止于转座子内部，偶尔也能转录到宿主DNA 内。甚至这种连续转录会激活毗邻的细菌基因，两种转录产物 IN RNA 和 OUT RNA 存在 40 bp 的重叠。从多拷贝抑制现象说明：IS 10 R 转座酶基因表达受 IS 10 R 元件的调控。在多拷贝质粒中导入额外的 IS 10 R 拷贝时，细菌染色体上 Tn 10 元件的转座作用便会下降。这种抑制效应需要 P_{OUT}启动子参加，而且是在翻译水平上起作用；由于从 P_{IN}和 P_{OUT}两个启动子产生的转录物，其 5'-末端都处于 40 bp 重叠区域内，OUT RNA 是含有 69 bp 的转录物。P_{OUT}启动子要比 P_{IN}启动子强得多，而用 OUT RNA 比 IN RNA 更为稳定，所以，OUT RNA 的水平比 IN RNA 高 100 多倍。实际上 RNA_{OUT}行使反义 RNA 的功能。因为这两种 RNA 的 5′ 端的重叠而能配对，过剩的 OUT RNA 促使它在 IN RNA 与核糖体连接之前早已迅速地与 IN RNA 结合，配对后的 IN RNA 不能作为合成蛋白质的模板，因此，不能被翻译，即 IS10R 转座酶基因不表达。

除了 OUT/IN RNA 的配对抑制转座酶合成外，单个转座子的转座作用会因复制后发生的甲基化作用而受到限制。

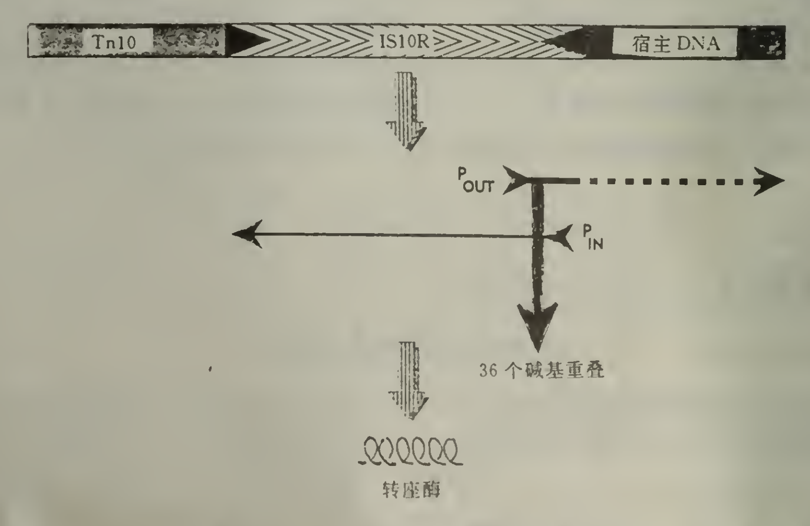


图 6-31　Tn10 中 IS10R 部分的组织结构

Tn10 和另一些转座子的转录酶启动子含有 GATC 序列，在大肠杆菌中这一序列的 A 被 Dam 甲基化酶催化甲基化了；因为GATC 的互补链也是 GATC(都从 5′→3′阅读)，此处 DNA 的两条链都被甲基化。在这种情况下，启动子便相对地钝化。但在 DNA 复制时，新合成的 GATC 中的 A 不立即被甲基化(可能有 5min 或更长)，这样，在新复制的两个拷贝中，只有一条链上甲基化，而另一条链上没有甲基化的启动子，比上述两条链都被甲基化的启动子的活性高得多，于是在 DNA 刚复制后出现少数的转座酶。与此类似，在转座酶的作用位点上，也就是在 IS10R 的末端反向重复序列内部也含有GATC，新复制的 DNA 是转座酶的最好的底物。因此，Tn10 在 DNA 刚复制后易发生转移。总之，甲基化作用阻止转座酶的合成，并且阻止转座酶结合到 DNA 上，从而限制转座，使转座频率降低。

综上所述，控制转座子转座频率的核心因素就是转座酶的水平，转座酶的数量通常是相当关键的。

五、转座子的某些遗传学效应

尽管原则上转座子在基因组内移动的概率相当小，但转座子作为位于基因组内的一个独立实体，它的不连续的活动似乎提供了某种自然选择。

转座子与基因组的关系就像寄生物与宿主的关系一样。如果转座会使必需的基因失活，或者转座子的数量过多，它就会成为细胞系统的负担，结果可能通过不利于转座发生的过程来平衡；但任何提供有利于选择的转座作用，如遗传重排，都将导致携带活性转座子的基因组优先存活。

转座因子的共同现象以及它们带来的遗传学效应有以下几个方面：

(一)引起插入突变

以 10^{-3}～10^{-8}频率进行的转座会引起插入突变。各种 IS，Tn 和 Mu 噬菌体都可以引起插入突变，如果插入的位置是一个操纵子的前端的基因，那么所造成的便可能是一个极性突变，极性突变是指减低蛋白质合成速度的基因突变。一般碱基置换突变没有极性效应，只有终止密码子突变和移码突变才有极性效应；由 IS1 和 IS2 的碱基序列分析中发现，它们确实存在无义密码子，当信息转移到 mRNA 中相应的无义密码子处，肽链的延长就终止，产生的是无酶活性的多肽片段，并且影响同一操纵子中其后基因产物的合成。IS2 以任何方向插入都有极性效应；IS3 只有以一个方向插入时才有极性效应。然而也有关于因 IS2 的插入而出现新的启动子信号的报道。

(二)插入位置上出现新的基因

如果转座子上带有抗药性基因，那么它一方面造成一个基因插入突变；另一方面在这一位置上出现一个新的抗药性基因。

(三)造成插入位置上出现靶 DNA 的少数核苷酸对的重复

不同的转座子插入后造成的受体靶位点处 DNA 的重复的碱基对数多少不等。如 IS1，Tn10 等造成 9 bp 的重复，IS3 造成 3 或 4 bp 的重复，IS4 造成 11 bp 重复。

(四)转座后原来位置上保持原有的转座子

插入序列和转座子从一个位点上转移到另一个位点上，原来位点上的这些结构依然存在，转座子实际上是以它的一个复制品转移到靶位点处的。

(五)引起染色体的畸变

尽管转座作用与细胞重组系统有差异，然而转座子作为细胞重组系统的底物，执行着携带同源区段的作用。在不同位点(甚至不同染色体上)的同一个转座子的两个拷贝可以提供相互重组的位点，这种交换导致缺失、倒位、插入或转座。

宿主 DNA 的重排可能由转座子的一个拷贝插入到邻近位置的第二个位点引起。宿主系统可以在两个拷贝的转座子之间产生重组作用，其结果取决于重复序列的取向。如果转座子的两个拷贝取向相同，则重组后产生缺失；如果取向相反，则发生倒位。

(六)切除

转座子可以从原来位置上消失，这一过程称为切除。切除的方式有两种：准确的切除和不准确的切除。准确切除的结果使插入突变的基因发生回复突变；如果这个转座子是一个带有抗药性基因的转座子，那么抗药性同时消失。不准确切割的结果留下转座子的残迹，而发生了插

入突变基因的不回复突变，但转座子本身所带的遗传标志消失，这是由不准确切除引起的染色体畸变造成的。这些现象均已在 Tn10 插入 *hisG* 的突变株的回复突变株中得到了证据。

在鼠伤寒沙门氏菌中有一个由 Tn10 插入而造成的组氨酸基因 *hisG* 的插入突变。Tn10 带有四环素抗性基因，所以，*hisG* :: Tn10（这表示 Tn10 插入到 *hisG*）插入突变型既是组氨酸缺陷型又是四环素抗性突变型 Tc^R。

自发的四环素敏感回复子以 10^{-5}～10^{-4}的频率发生。在 Tc^S 回复子中一部分是原养型的，它们是 Tn10 准确切除的结果；一部分 Tc^S 菌株仍然是组氨酸缺陷型，它们可发生倒位或缺失，这是不准确切除的结果。

总之，由转座子介导的基因重排在进化中非常重要。例如，转座子可将一些原来在染色体上相距甚远的基因组合到一起而形成一些操纵子；也有可能把两段分离着的碱基序列转变成连在一起的DNA。这样在进化过程中就产生一些新的蛋白质。在细胞繁殖的一个世代中，基因重排可能也是非常重要的。例如，转座子可以看作是一个开关，在一个特殊的位点发生插入作用或倒转一个含有启动子的序列，就可以使基因打开或关闭。

六、真核生物的转座子

转座子首先在玉米中发现，其它真核生物中的转座子则直到 60 年代后期，在大肠杆菌的半乳糖操纵子的研究中第一次发现插入序列后才陆续发现的。真核生物中转座子的转座机制有的与细菌转座子相似，其遗传信息的流动从 DNA 到 DNA，如玉米的 Ac-Ds 元件，果蝇 P 元件和 FB 元件等。有的则与逆转录病毒相似，其遗传信息从 RNA 到 DNA 再到 RNA，如逆转录病毒、果蝇的 copia 元件和酵母的 Ty 元件等。

（一）玉米的控制元件

早在本世纪初，玉米遗传学家已经鉴别了在玉米中影响籽粒色素的不稳定突变，第一个发现转座子的就是玉米中的控制元件(controlling element)。Barbara McClintock 从 1942 年到冷泉港实验室后就开始了玉米控制元件的研究。1951 年发表了关于玉米控制元件的论文。她的开创性工作，打破了孟德尔关于基因在各个独立的染色体上固定排列的概念。控制元件没有固定的染色体定位，它们能抑制与其毗邻的其它玉米基因的表达，即控制元件在一个部位上的插入可影响邻接基因活性，使基因有较高的突变频率；并且控制元件出现的部位可以引起缺失、复制、倒位和易位。

在玉米基因组中含有好几组控制元件的家族。在每种玉米的品系中，它的数目、类型和插入部位都有其特征。同一家族的成员又可以分为两类：即自主元件和非自主元件。自主元件(autonomous element)，具有切割和转座能力。因为它的持续活性，在任何位点上的插入会产生一个不稳的或易变的等位基因。自主元件本身的缺失或其转座能力的缺失，可把易突变的等位基因变成稳定的等位基因。非自主元件(nonautonomous element)，它们不能转座或自发地经受其它变化。只有当同一家族的自主元件在基因组的某处存在时，非自主元件才变成不稳定的。而当非自主元件与自主元件反式互补时，才会显示出与自主元件相联系的普遍活性，包括转座到新位点的能力。自主元件转座作用所需要的反式功能缺失时就产生自主元件。

玉米控制元件中的每一个家族是由一个简单类型的自主元件和多种类型的非自主元件组成。非自主元件之所以列入此家族是因为它可以被自主元件反式激活。

玉米中转座子具有共同的结构特征。其末端具有反向重复及邻接的靶 DNA 的短的正向

重复，但在大小和密码的容纳能力方面有所不同。对自主元件和非自主元件的 Ac-Ds 家族的研究已最为清楚。在玉米转座子研究中最初所用的品系的第九染色体上有一个解离子(dissociate，Ds)，在另一染色体上有活化子(activator，Ac)，构成 Ac-Ds 系统。Ac 元件是一个完整的转座子，其组织结构与细菌转座子非常相似，长度为 4 563bp，转录后的单链 RNA 经剪接后形成 3 500bp 的 mRNA，其中有编码 801 个密码子的序列，与它含有 5 个外显子的单基因表达相一致，此序列占 Ac 的长度的绝大部分。在编码序列的两侧带有 11bp 的短反向重复；在插入位点上复制了 8 bp 的靶序列。因为 Ac 具有活性的转座酶基因，所以，使 Ac 可自主地移动。Ac 元件能引起较高回复频率的突变。基因中的 Ac 突变可以合成玉米粒色素花色素苷，因此，在发生回复突变时，会在正常籽粒颜色的背景上产生带有杂色的斑点。

Ds 元件的长度和序列是可变的，但都与 Ac 有关。实际上，Ds 元件是在转座酶基因上有缺失的 Ac 元件，其缺失的长度变化不定，在 Ds 9 中仅缺失 194 bp，Ds 6 缺失长达 2 kb。Ds 元件比 Ac 元件短，但它同样有 11 bp 的反向重复末端。Ds 本身不能移动，Ds 元件必须有 Ac 元件(Ac element)的存在才起作用。Ds 元件能引起稳定的突变，这些突变产生无色的玉米粒(除非这些植物与 Ac 品系杂交)。当 Ac 存在并且具有转座酶的活性时，Ds 就能像复合转座子那样跳跃并活化花色素苷基因。虽然 Ds 和 Ac 不在同一条染色体上，但 Ac 元件通过所产生的某些可扩散的物质作用于 Ds 元件而影响 Ds 元件(Ds element)，Ds 元件作用于染色体，使之发生断裂。Ac 元件越少，则由 Ds 引起的染色体断裂得越早，颜色斑点越大。

Ac 激活 Ds 而引起染色体断裂的结果以一个杂合子为例，Ds 位于一条同源染色体的着丝点和一系列显性标记之间；另一条同源染色体无 Ds 而带有隐性标记 *Cc*，*bz*，*wx* 的无着丝点片段。Ds 处断裂产生一个带有显性标记的无着丝点片段，因为该片段没有着丝点，所以，在有丝分裂中丢失，导致表现型的改变。所以，其后代细胞只表现由完整染色体携带的隐性标记所决定的表现型。玉米体细胞发育时出现的表型改变称斑驳现象(variegation)，表现为一块新表型的扇状体嵌合在原表型组织内。

(二)果蝇中的 P 元件

黑腹果蝇的某些品系杂交时，其后代显示了不育特征。一系列的突变、畸变破坏了减数分裂，结果导致不孕，这些缺陷现象称为杂种不育(hybrid dysgenesis)。

当果蝇的 P 型(paternal contributing)和 M 型(maternal contributing)的两个品系杂交时，形成不对称系统，即当 P 雄蝇与 M 型雌蝇杂交时产生的杂种后代是不育的，而反交则正常。不育原则上是一种种系生殖细胞现象，在 P-M 杂交系统中，F_1 杂种果蝇有正常体组织，但它们的性腺不发育，配子中发育形态缺陷始于种系生殖细胞快速分裂阶段。

在与 M 型雌蝇杂交时，P 雄性的任何一个染色体都能诱发不育。重组染色体显示在每个 P 品系染色体中有若干区段能引起不育。这表明 P 雄性有大量的 P 因子(P factor)，其序列分布在许多不同的部位上，不同的 P 品系中 P 因子的位置不同，M 品系果蝇染色体上没有 P 因子。由杂种不育的突变型 DNA 作图得知：所有突变是 DNA 序列的插入造成的，该插入序列称为 P 元件。P 元件的插入形成了一个典型的转座系统。可见杂种不育的原因可能是果蝇的 P 转座元件转座引起的结果。单个 P 元件的长度是变化的，但在序列上是同源的。携带 P 元件的果蝇品系通常只携带几个全长转座子，最长的 P 元件长约 2.5 kb，有 4 个转座作用都需要的可阅读框架(open reading frame，ORF)，推测它们所编码的转座酶和阻遏物可能与细菌转座子移动所需要的一样。大多数 P 序列像 Ds 一样，是由全长的完整 P 元件向内部缺失产生的衍生

物。这些元件至少丧失了产生转座酶的能力，因此，它们要依赖全长元件编码转座酶才能移动。缺失的 P 元件衍生物的大小为 0.5～1.4 kb，两侧边界都有 31 bp 的末端反向重复，P 元件插入时形成了一个典型的转座系统，在靶 DNA 插入位点处产生 8 bp 的正向重复 DNA 序列。一个 P 品系带有 P 元件的拷贝数是 30～50 个，其中 1/3 是全长的，但 M 品系中无 P 元件。在 P 品系中，P 元件就像染色体上的惰性组分，只有当 P 雄蝇与 M 雌蝇杂交时，它们才被激活。即由于进入的雄性 DNA 编码的转座酶迅速产生，引起 P 元件移动，而且通过新的 P 元件的插入引起频繁的突变，这种劣生杂交的子代表现出较低的生育能力和染色体的畸变。P 元件的最初转录产物长为 2.5～3.0 kb，长度的差异可能只反映了终止位点的渗漏，转录产物合成两种蛋白。P 元件的活性具有组织专一性，它只存在于种系中，但在种系和体细胞组织中都能转录，组织专一性是通过改变剪接方式实现的。在体细胞组织中，仅前两个内含子被剪切，即外显子 2，外显子 3 不发生剪接，产生的 mRNA 组成是 ORF_0-ORF_1-ORF_2，其翻译产物是 66 kD 的蛋白质，该蛋白质是转座子活性的抑制物，起阻遏作用，而不是启动转座作用的发生。这种调节形式就限制了 P 元件转座到生殖细胞中。

在种系组织中，另一种剪接反应去除了第 3 个内含子，这样就全部 4 个外显子剪接到一起，得到的 mRNA 翻译产物是 87 kD 的蛋白质，即转座酶。可见 P 元件转座通过可变剪接机制只限于在果蝇的种系生殖细胞中，活性转座酶由 4 个外显子编码，而最后的剪接则只发生在生殖细胞中，负责这种组织特异性剪接的体细胞中含有一种蛋白质，它可以结合到外显子 Ⅲ 的序列上，从而阻止了最后一个内含子的剪接。在种系生殖细胞中，由于这种蛋白质的缺失可以产生编码转座酶的 mRNA。如果用基因工程手段把一个 P 元件的最后一个内含子准确切除，则第三、四外显子就能像剪接一样连在一起。这样经人工改造的 P 元件在果蝇的体细胞中很容易地进行转座。

P 元件的转座需要 150 bp 长的末端 DNA，转座酶结合于 10 bp 上，后者与 31 bp 的反向重复邻接。转座发生的机理类似于 Tn10，通过非复制型的切割和粘连进行转座。已转座的元件插入到新位点可以引起突变；留在供体部位的断裂具有有害的影响，结果导致产生杂种不育的后代。

M 品系的雄性与 P 品系的雌性杂交或 P×P 杂交都不会劣生，因为当完整的 P 元件存在于雌性基因组时抑制了许多短 P 元件的转座作用，卵的细胞质可能早已充满了阻遏物型分子。杂种不育对杂交时性别定向的相依性表明了细胞质和 P 因子本身同样重要，因此，把细胞质的作用称为细胞质型(cytotype)。含 P 元件的果蝇品系中有 P 细胞质型，而无 P 元件的果蝇品系有 M 细胞质型。

细胞质型的效应可做分子水平的解释。它藉助于 66 kD 蛋白来抑制转座。该蛋白在卵中作为母系因子。在一个 P 家族中，必须有足够的阻遏蛋白才能阻遏转座的发生，即使 P 元件存在也是如此，在任何有 P 元件雌性杂交中，该 66 kD 蛋白存在时或阻止转座酶的合成或抑制转座酶的活性。但当雌性亲本是 M 型时，在卵中不存在阻遏物，从雄性亲本导入 P 元件时生殖品系上则产生了转座酶活性，P 细胞质型至少在一个世代后才能发挥作用，因为它必须经过下一世代的成虫卵中有充足和稳定的阻遏蛋白。

P 元件转座引起的杂种不育和其它转座体系相类似，在物种形成过程中起重要作用。因为杂种不育降低了近交能力，所以，它是物种进化中的一个阶段。假如某一地区通过转座子产生一种不育系统，而在另一地区由于其它原因产生不同的不育系统，这两个地区的果蝇将是两个系统间杂种不育。如果这使得它们之间杂交不孕，那么两个群体就成为遗传性隔离。因而多重

不育系统的进一步出现将最终导致不能交配，即物种形成。此外，最近P单元已成为将基因导入果蝇基因组新位点上的有效载体。

（三）逆转录转座子

逆转录病毒能将其RNA基因组的DNA拷贝（原病毒）插入宿主细胞的染色体上而发生转座作用，因而称之为逆转录转座子（retrotransposon），亦称逆转录子（retroposon）。目前已知，在所有高等生物的基因组构成中，也存在着与逆转录病毒基因组非常相似的逆转录转座子，如酵母的Ty转座子和果蝇的copia转座子，它们位于较长的正向重复或长末端重复（long terminal repeat，LTR）区域内，在转录、逆转录及整合过程中发生移动。

逆转录转座子具有着一些与简单转座子相似的特性：

（1）可能与高频率的自发突变有关。

（2）刺激寄主细胞基因组发生多种形式的遗传重排，如相邻细胞DNA的缺失、倒位、加倍和易位。

（3）插入单元两端的DNA序列在插入过程中可能稍有增加（3～13 bp）。

（4）转座单元的末端含有长度为2～50 bp的反向重复。

（5）转座过程经常与转座单元的复制同时进行。这一点对逆转录转座子来说是必需的，在原核生物的转座中则时有发生。

（6）转座子通过编码蛋白质，以顺式或反式方式作用于转座单元，来对自身的转座功能进行调控。

就逆转录转座子而言，逆转录病毒与真核生物的酵母Ty转座子、果蝇copia转座子之间的差别在于逆转录病毒中存在另外一个基因，也就是编码逆转录病毒外层类脂包膜糖蛋白的*eny*基因。逆转录病毒是一种真正的二倍体病毒。其粒子中除了含有两个基因组RNA拷贝以外，还含有一种在基因组逆转录合成DNA过程中作为引物的细胞tRNA。

1. 逆转录病毒基因组　逆转录病毒基因组编码逆转录酶和整合酶，逆转录酶在侵染细胞的细胞质中把RNA转换成线型双链DNA，线型DNA进入细胞核，由整合酶催化整合作用，整合后的原病毒DNA通过宿主系统产生病毒RNA，所以，它们以mRNA作为基因组包装进病毒颗粒中，整合至感染细胞的逆转录病毒序列作为内源病毒留在细胞基因组中。

（1）基因组结构　典型的逆转录病毒基因组含有3个或4个基因。具有3个基因*gag-pol-env*序列的结构。它们表达成熟产物的功能如下：*gag*基因：产生病毒颗粒核蛋白核心的蛋白质部分，即DNA结合蛋白。*pol*基因：编码与核酸合成和重组有关的多肽复合物，它具有3种功能不同的活性，即逆转录酶、整合酶和RNA酶H的活性。*env*基因：编码病毒颗粒被膜的组分。

逆转录病毒基因组两端还有很长的非编码序列，在5′端有帽子结构，在3′端有聚腺苷酸尾，除了这两个转录后加工的产物外，基因组本身在两端有完全相同的10～80碱基正向重复序列R。其长度因病毒株系而有所变化。这两个R片段的内侧，5′端部分为80～100个碱基长的U5，3′端部分为170～1 350个碱基长的U3。在U5的内侧为引物结合位点（primer binding site，PB）该引物是RNA基因组复制时从宿主细胞中得来的tRNA。在不同的逆转录病毒中tRNA的种类也不同。这些tRNA都是通过3′端部分与PB位点互补，以便在侵入细胞后引发DNA的合成。

逆转录病毒RNA、线型DNA和原病毒DNA的结构：①病毒RNA的两端是正向重复R；②病毒线型DNA两端为LTR；③原病毒DNA两端是缩减2 bp的LTR。

（2）逆转录病毒RNA转换为原病毒DNA　从病毒RNA逆转录为病毒线型DNA，并非RNA基因组的精确拷贝，而是在5′端增加了U3，在3′端增加了U5，产生了两个完全相同的正向重复序列U3RU5，称为长末端重复LTR。在长末端重复中存在转录起始信号和3′-端切割并加聚腺苷酸尾的信号。在逆转录过程中，逆转录酶以tRNA为引物，以U5R为模板，产生大约100～200核苷酸的cDNA，当逆转录酶到达基因组5′端时，该酶的RNA酶H活力切除刚才拷贝过的模板，于是新生的DNA序列U′5R′通过R′与R的互补关系与3′端的R序列相结合（第一次跳跃），因此产生了新的模板——引物配对。利用新的模板，逆转录酶能产生全部RNA的互补链。为了转变成双链DNA，在RNA模板U3序列的左边有一个切割位点，从这里切断RNA，并以其3′-OH为引物，开始了正链DNA的合成，产生了正链DNA的U3RU5PB，然后出现了第二次跳跃，即tRNA被RNA酶H活性所分解，而使正链上的PB与负链上的PB′互补，RNA酶H又降解了整个基因组RNA，最后将两条DNA链的合成进行到底，产生两端具有U3RU5的线状DNA。

整合的原病毒DNA其结构与线型DNA相似，LTR两端是反向重复，在其两侧则是靶DNA的正向重复。现已证明逆转录病毒是以线型DNA的形式进行整合的。在整合过程中，病毒DNA的末端至关重要。在紧靠每个LTR反向重复的末端存在两个保守的碱基CA，整合酶覆盖型状DNA的两个末端，切除保守的CA之外的碱基，把平齐末端变成凹陷末端。同时整合酶随机选取宿主DNA靶位点进行交错切割，两个切口由4或6 bp分隔。整合酶催化由靶位点断裂所产生的5′端，与病毒DNA3′-凹陷端共价相连，而病毒DNA凸出的5′端则由宿主细胞的酶修复。所以，整合后的原病毒DNA在LTR中丢失2 bp，相当于5′端U3左边2 bp和3′端U5右边的2 bp。

所有原病毒DNA上与细胞基因组DNA相连接的两端序列都是5′-TG……CA-3′，而且酵母Ty元件，果蝇的copia元件以及Mu噬菌体的两端也都是同样的序列，可见它们在进化过程中有着密切的关系。

2. copia元件　copia元件（copia element）是果蝇中另一种转座元件，它有几个家族（多于11个），每个家族的成员都相当保守。在果蝇基因组中，copia元件的拷贝数为20～60个，取决于果蝇的品系，这些DNA序列重复移动，并能在基因组内随机分布。copia元件的插入能引起若干基因的稳定突变。copia的长度约为5 000 bp，两端具有相同的276 bp的正向长末端重复，其长度和序列在家族之间有所变化，每个末端各有13 bp的不完整的反向重复。与其它转座子一样，它插入到一个新位点，在靶DNA上产生5 bp的正向重复的复制。

最近发现copia元件本身可以游离环型DNA分子的形式单独存在，copia环状分子有两种，一种是5 000 bp；另一种缺少276 bp的末端重复，约为4 700 bp。就像逆转录病毒的DNA环一样，copia元件的一个特有的性质是它们以高度转录其全长并产生聚腺苷化的RNA，这样的转录仅限于细胞核。copia序列只含有一个较长的阅读框架，它与逆转录病毒的*gag*和*pol*序列间存在同源性，因此，亦同源于逆转录酶、整合酶和核内酸性结合蛋白，但与包装病毒所需的逆转录病毒的*env*序列就没有同源性，这说明copia不可能产生病毒样粒子。

3. 酵母中的Ty单元　Ty是酵母转座子（transposon yeast）的缩写。Ty单元插入到编码序列或调节序列时，除了有灭活基因的功能外，有时Ty的插入还能使沉默的基因激活。这种行为已由克隆和序列分析得到证实。

酵母基因组中存在的Ty元件有两种重要的类型，Ty1和Ty17，它们的结构大致相同。Ty 1元件长6.3 kb，两端各有334 bp的正向重复，称为δ序列，它们也被称为LTRs。重复序列含

有 RNA 聚合酶Ⅱ的强启动子，导致完整元件从左边 δ 序列转录成 5.6 kb 的 mRNA，从右边 δ 序列转录宿主的 DNA，这就是 Ty1 元件能激活沉默基因的原因。Ty 元件携带两个开放读框，它们同方向表达但不同步阅读，并重叠 13 个氨基酸。TyA 序列编码 DNA 结合蛋白，*TyB* 基因编码一个长的多聚蛋白，加工后成为几个不同的酶，依次为蛋白酶、整合酶、逆转录酶和 RNA 酶 H，以 Ty mRNA 逆转录为 DNA，并且把它整合到宿主基因组。这个基因在动物中与逆转录病毒的 *pol* 基因有关。

Ty 转座子与前述的果蝇 copia 元件相类似，它也与在脊椎动物细胞中增殖的反转录病毒基因组有很多相似的性质。当它插入逆转录病毒基因组能激活邻近基因的表达，从而产生肿瘤。

Ty 元件和 copia 元件的结构不同于细菌转座子，它们的转座机制也不同于细菌转座子（通过信息从 DNA 传递到 DNA）。Ty 元件通过 RNA 中间体转座，这是利用重组 DNA 技术，在人工构建的 Ty 基因组的一系列实验中被发现的。人工构建的 Ty1 衍生物是可被半乳糖诱导的酵母 *GAL*1 基因的启动子替代左侧 δ 中的启动子，且插入的 GAL1 片段是第一个内含子的酵母基因。工程构建的 Ty 转入酵母细胞后可在葡萄糖或半乳糖上生长。当细胞生长在半乳糖上，人工 Ty 元件以高水平转录，这就提高了它们的转座频率；而且这个带有内含子的人工 Ty 转座后的单元中精确地缺少了内含子序列。因为去除内含子的方法只有一种，即 RNA 剪接，所以，新的转座子元件一定是由供体元件转录的剪接 RNA 分子的逆转录产物——DNA 拷贝。因此，酵母的 Ty1 转座子也被称为逆转录转座子。虽然 Ty 质粒并不产生感染粒子，但在诱导产生转座的细胞内部却积累病毒样颗粒，其中含有 RNA 的全部长度，双螺旋 DNA，具有逆转录酶和整合酶活力的 TyB 的产物。TyA 产物就如 *gag* 前体一样断裂后产生“VLP”的成熟的核心蛋白，这使 Ty 转座子与逆转录病毒之间的相似性更进了一步，因此，Ty 的转座子就像丢失了 *env* 基因而不能包装其基因组的逆转录病毒。

参考文献

1. 弗雷费尔德.D.分子生物学(下册).蔡武城等译.北京：科学出版社，1991.第 18 章，p.142～167，第 19 章，p.167～189.
2. Watson，J. D.，Hopkins，N. H.，Roberts，J. W.，Steitz，J. A.，Weiner，A. M. Molecular Biology of The Gene. I(4th ed.). Benjamin/Cummings，Menlo Park：1987. p. 313～338，p. 663～666.
3. Watson，J. D. et al. Molecular Biology of The Gene. II(4th ed.). Benjamin/Vummings，Menlo Park：1987. Chapters23，p. 850～871，Chapters24，p. 936～934.
4. Lewin，B. Genes V. Oxford University Press，Inc. New York：First Published 1994. Chapters33，p. 967～997，Chapters34，p. 999～1 031，Chapters35，p. 1 033～1 056，Chapters36，p. 1 059～1 086，Chapters37，p. 1 108～1 130.
5. Stryer，L. Biochemistry. (4th ed.). W. H. Freeman and Company，San Francisoo：1985. p. 368～378.
6. Lehninger，A. L.，Nelson，D. L.，Cox，M. M. Principles of Biochemistry. (2nd ed.). Worth Publishers，Inc. New York：1993. p. 839～852.
7. Lodish，H.，Baltimore，D.，Berk，A.，Zipursky，S. L.，Matsudaira，P.，Darnell J. Molecular Cell Biology (3rd ed.). W. H. Freeman and New York：1995. p. 320～342.
8. Kornberg，A. & Barker，T. A. DNA Replication (2th ed.). W. H. Freeman and Company，1992. p. 792～822.

（曹淡君）

第七章 RNA 的合成

第一节 引　　言

细胞中的遗传信息是从脱氧核糖核酸(DNA)到核糖核酸(ribonucleic acid,简称 RNA),再由 RNA 到蛋白质。第五章已讲过 DNA 的合成(复制),我们在本章讨论 RNA 的合成,RNA 的合成是以 DNA 为模板,合成核糖核酸链,这个过程称为转录(transcription)。虽然 RNA 的合成和 DNA 的合成,从核苷酸的水平上看来,是很相似的,但此二过程的生物学意义却十分不同。DNA 的合成仅是基因组的精确复制,并且固定不变。但基因组转录成 RNA,却涉及许多遗传和表达的复杂性。例如,基因通常在转录时要进行调节,RNA 在合成的不同时期会受到改变:一条 RNA 链在 DNA 模板上的转录,从何处起始,何处停止,何时起始,何时停止,如何使二者达到协调一致等,都是很重要的问题。有关基因表达的分子生物学最新进展,都是来自对 DNA 序列和控制转录的蛋白质因子的研究,特别是对 RNA 聚合酶的研究。第一次积累的基本资料是从细菌(主要是大肠杆菌)RNA 聚合酶得来的,这些研究结果所建立起来的原则在原核生物中具有普遍意义,但在真核生物中就要复杂得多。本章将从 RNA 的基本知识谈到原核生物的 RNA,并进一步谈到真核生物的 RNA。

一般说来,DNA 分子所贮存的蛋白质遗传信息,必须转变成信使 RNA 分子(mRNA),才能到达蛋白质合成的工厂即核糖体,然后,蛋白质合成酶系才能把 mRNA 所带来的信息翻译成蛋白质。在合成蛋白质过程中,还需要有能携带氨基酸的 tRNA 和构成核糖体的 rRNA。这 3 类 RNA 都必须以 DNA 为模板,在依赖于 DNA 的 RNA 聚合酶催化之下合成,包括 RNA 链的起始、延伸和终止等环节,上述一系列过程就叫做转录或 RNA 的合成。转录是基因表达的第一步,也是关键的一步,其所以是关键,就在于生物体是否决定让该基因转录,这种“决定”的调控机制也是极其重要的。

转录过程有两个重要方面:一是 RNA 合成的酶学过程;二是 RNA 合成的起始和终止及其调控。早在 50 年代,人们就对 RNA 的合成酶进行了研究,了解比较清楚;而对 DNA 分子上的特定序列,其研究的开展还是近 10 多年来的事,在 DNA 重组技术和序列分析发展的情况下,DNA 分子上的特定序列研究,也获得了成果。本章中对 DNA 的两条链采取了与 60 年代相反的新命名法,即把不做模板的 DNA 链称为有义链(sense strand),又称编码链(coding strand);把做模板的链称为反义链(antisense strand)或模板链(template strand),这是新的表示方法,因为非模板链的核苷酸序列与转录出来的 RNA 序列一致,只是 T 与 U 有所区别。对于 mRNA,也易于将 DNA 序列直接与氨基酸密码子联系起来,按照新的表示方法,如 ΦX 174,其 DNA 进入宿主细胞后必须复制出一条互补链而成为双链超螺旋的形式,才能从这条互补链上转录出它所需要的 RNA。这样,这条互补链就称为反义链,又称为负链,而能够进入噬菌体颗粒的 DNA 链则称为有义链,又称正链。对于单链 RNA 病毒,若其基因组 RNA 能

够作为 mRNA，则该 RNA 病毒就称为正链 RNA 病毒；若基因组 RNA 进入宿主细胞后必须通过 RNA 复制，产生出互补链作为 mRNA，则这种病毒就称为负链 RNA 病毒。

第二节　原核生物 RNA 的结构及种类

一、RNA 的结构

（一）RNA 的化学结构

RNA 合成的前体是 ATP、GTP、CTP 和 UTP 等 4 种 5′-核苷三磷酸(rNTP)。每个 NTP 的核糖部分有两个羟基，各位于 2′和 3′-碳原子上(图 7-1)。

在聚合反应中，一个核苷酸的 3′-OH 基团与第二个核苷酸的 5′-磷酸基团发生反应，释放出焦磷酸，形成磷酸二酯键(图 7-2)。这与 DNA 合成中发生的反应相同。

如图 7-2 所示，DNA 模板链上的 C，T，G，A 分别与 RNA 分子中的 G，A，C，U 配对，被转录的 DNA 分子虽系双链，但只有一条链做模板。RNA 链按 5′→3′方向延伸，与模板 DNA 链呈反平行。

图 7-1　5′-核苷三磷酸

2′-OH 基团用黑体字表示，
在脱氧核糖中被 H 所取代。

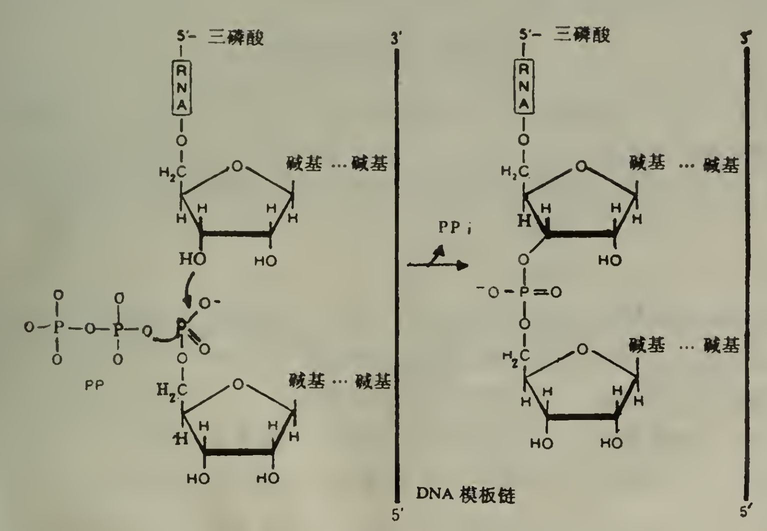


图 7-2　RNA 聚合酶催化链伸长的反应机制

焦磷酸基团(PPi)和框中的氢原子不出现在 RNA 链中。DNA 模板链和
RNA 链是反平行的，和双链 DNA 中的情况一样。粗箭头连接反应基团。

（二）RNA 形成大分子后的结构

与 DNA 不同，RNA 几乎常常是单链，它以各种方式在 DNA 模板上合成。有的 RNA 分子像蛋白质那样，具有确定的结构，有的 RNA 的结构则是不太确定也不十分稳定。几乎每个 RNA 分子都有许多短的双螺旋区域，双螺旋区域的形成是因为在一条 RNA 链的发夹结构之内碱基对以反平行方向靠氢键结合之故(见图 7-3)。除了正规的 AU 和 GC 碱基对之外，结合得较弱的 GU 碱基对在形成 RNA 结构中也起作用。还有些 RNA，如 tRNA，还含有较多的互

补序列，互补序列使整个 RNA 分子形成特殊形状，如三叶草结构。二级结构也在单链 RNA 中存在，如形成一条长的双螺旋结构等(见图 7-3)。

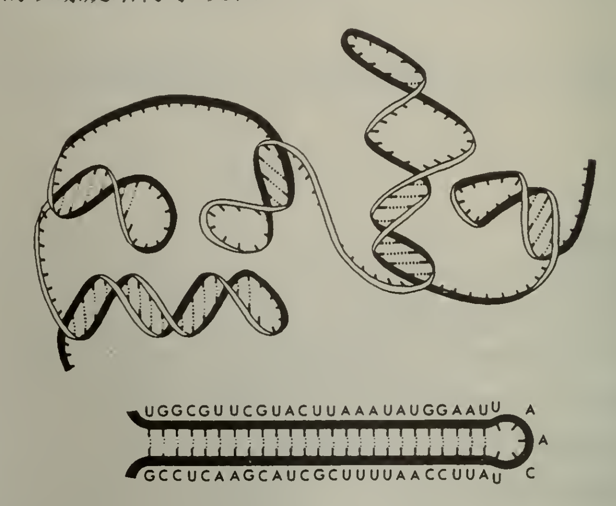


图 7-3 RNA 的某些结构

每个细胞中均含有许多不同的 RNA 分子，其长度从小于 50 个核苷酸大到万个以上的核苷酸，几乎都是一些线状的单链，也有一些为环状的 RNA 分子。

二、RNA 分子的种类

RNA 分子有 3 大类，即信使 RNA(messenger RNA 简写 mRNA)、核糖体 RNA(ribosomal RNA 简写 rRNA)和转移 RNA(transfer RNA 简写 tRNA)。所有这些 RNA 都按 DNA 碱基序列合成。这 3 类 RNA 的结构与功能均不一样，原核与真核的 RNA 也不一样，此处先谈原核的 RNA。

(一)mRNA

把 DNA 模板链上的碱基序列，转录为 RNA 分子上的碱基序列(mRNA)，再从 mRNA 上的碱基序列通过合成蛋白质的机构，获得氨基酸的序列，即为 mRNA 的功能。mRNA 的碱基序列从起始密码到终止密码以 3 个碱基为一组而读码，3 个一组的碱基称为密码子(codon)，每个密码子对应一个氨基酸或一个终止信号。对应于一条多肽链的 DNA 片段加上起始信号、终止信号称为顺反子(cistron)，只编码一条多肽链的 mRNA 称为单顺反子 mRNA(monocistronic mRNA)。由一条 mRNA 分子编码几条不同的多肽链也是常有的，这种 mRNA 称为多顺反子 mRNA(polycistronic mRNA)。多顺反子 mRNA 往往对应于一个代谢途径中的各种蛋白质。例如，在大肠杆菌中半乳糖代谢需要 3 种蛋白质，它们都是从一条mRNA分子合成的；组氨酸合成所需的 10 种酶是由另一条 mRNA 分子编码的。细胞利用多顺反子比较经济，因各种蛋白质的合成便于统一调控。mRNA 分子的大小不同，最小的有 150 个核苷酸长，一般为 900～1 800 个碱基，大的如组氨酸 mRNA，编码 10 个蛋白质，含有 12 000 个核苷酸。除了顺反子、转录所需的起始序列和终止序列外，mRNA 中还有位于编码区前的前导区(leader)，此区位于 mRNA 的 5′-P 端和起始位点(肽链合成)之间，有些前导区中含有一个调节区称为弱化子(attenuator)，它调节蛋白质合成的速度。在 5′-P 和 3′-OH 两个末端都发现过非翻译区，多顺反子 mRNA 分子还含有顺反子间序列，称隔离序列(spacer)，长达几百个碱基。mRNA 的寿

命较短，典型的 mRNA 分子半衰期是几分钟。

（二）核糖体 RNA

核糖体是合成蛋白质的场所，由核糖体 RNA 和具有各种功能的大量蛋白质组成。核糖体 RNA 共分 3 种，原核生物的由 5 S rRNA、16 S rRNA 和 23 S rRNA 所组成，分别含有 120，1 541和 2 904 个核苷酸。rRNA 是稳定的分子，其前体比成熟的 rRNA 分子大。

（三）转移 RNA

tRNA 是传递氨基酸的 RNA，此类 RNA 的分子也比原始转录物小得多，并且富含稀有碱基，稀有碱基都是转录后加工形成的，成熟的 tRNA 常常形成三叶草结构。

第三节 RNA 酶促合成的特点

一、RNA 在 DNA 模板上酶促合成

前已述及，RNA 实际上像 DNA 一样，是长的、不分枝的链，由 4 种不同的核苷酸直接将 DNA 链上的遗传信息转移到互补的 RNA 核苷酸序列上，腺嘌呤与 RNA 的尿嘧啶；鸟嘌呤与 RNA 的胞嘧啶配对。这说明，某些控制机制必须使 DNA 链分离，并为互补的 RNA 链充当模板。

上述推测为 RNA 聚合酶（RNA polymerase）的发现所证实。RNA 聚合酶存在于一切细胞中，一个大肠杆菌细胞大约有 3 000 个分子的 RNA 聚合酶，此酶将核糖核苷酸聚合在一起，催化形成核苷酸之间的 3′→5′磷酸二酯键，形成分子的骨架（见图 7-4）。酶以这种方式作用时，必须有 DNA 的存在，DNA 必须将正确的核苷酸排成一列，以便于 RNA 聚合酶起催化反应。

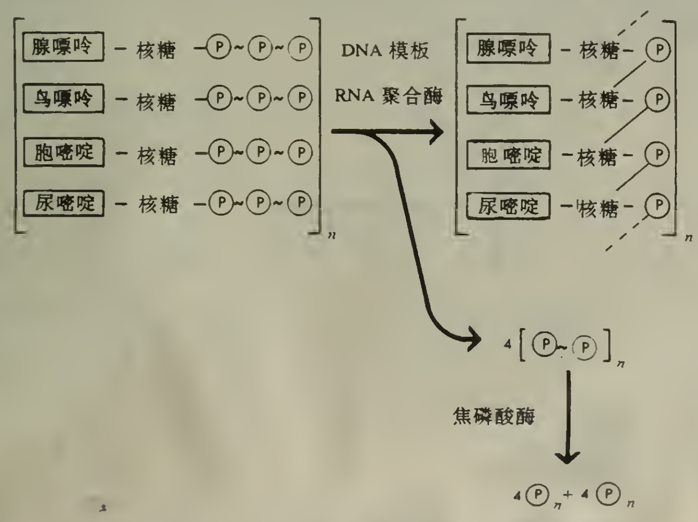


图 7-4　在 RNA 合成中催化形成的 3′→5′磷酸二酯键

虽然，RNA 聚合酶在合成 RNA 时通常用的是双链 DNA，实际上酶用的模板都是单链 DNA。当酶通过时，双螺旋 DNA 一小段解链，酶用单链 DNA 做模板合成 RNA。此现象在 ΦX 174 病毒的实验中观察得很清楚。此病毒属于单链 DNA 病毒，当其作为 RNA 的模板时，只有一条 DNA 链存在，而它的产物则是一条与 DNA 链互补的 RNA 链（见表 7-1）。这时可以分离出来 DNA-RNA 的杂交双链。当模板是双链时，RNA 合成后便迅速脱离 DNA 模板，两个分开

的 DNA 链又重新结合在一起，不能分离出 DNA-RNA 杂交链。

表 7-1 用单链 ΦX 174 病毒 DNA 做模板酶促合成 RNA 的碱基组成

	ΦX 174 DNA	RNA 产物的观察值	预期的 RNA 组成
腺嘌呤	0.25	0.32	0.33
尿嘧啶	0.33(胸腺嘧啶)	0.25	0.25
鸟嘌呤	0.24	0.20	0.18
胞嘧啶	0.18	0.23	0.24
总 计	1.00	1.00	1.00

由此，我们看到 RNA 的合成与 DNA 的合成原理基本上是一致的。这两种合成的直接前体都是核苷三磷酸，从它的一个焦磷酸键获得能量促使反应走向合成；另一种相似的情况是，两种合成都是一个酶为 4 种核苷酸工作，为了碱基的配对，必须做出正确的选择。

RNA 合成过程也很精确，但不像 DNA 有校对机制，因此，精确程度不及复制，DNA 的复制系统相当完美，拷贝得很精确，而细胞中的 RNA 不是自我复制的，所以，能产生非遗传上的差错。

二、只有一条 DNA 链作为一个基因的 RNA 模板

假如一个已知基因两条 DNA 链的每一条都可作为 RNA 的模板，则每个基因将会产生两个 RNA 产物，而此两个 RNA 产物是互补的。但是，遗传学的证据说明，每个基因只有一种蛋白质。这就有两种可能，一种是两条 DNA 互补链只产生出一条 RNA；另一种是合成了两个 RNA，但只有一条 RNA 有功能。看来前一种可能性是对的，在生物体内只发现一条 RNA，这在检测病毒 SP8 合成 RNA 的过程中就可以获得证据。SP8 病毒是在枯草杆菌(*Bacillus subtilis*)中繁殖的，而且碱基序列很独特，很容易与别的 RNA 或 DNA 的序列区分开，这样就易于鉴别其 DNA 是产生了一条 RNA 或两条 RNA。

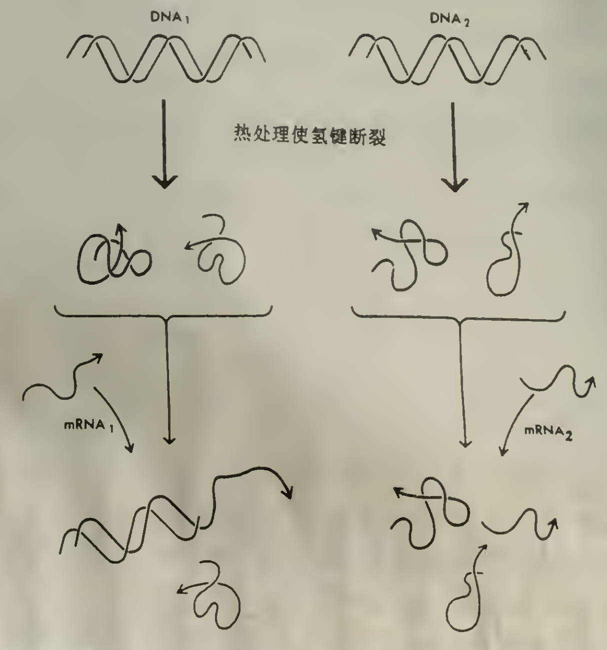


图 7-5 只有一条 DNA 链与 RNA 互补试验

此项试验的过程是，加热双链 DNA，使分开成单链，再与此 DNA 产生的 RNA 混合，使形成人工的 DNA-RNA 杂交分子，即可检测出结果。具体的做法是，加热 DNA 分子，温度要恰好低于 100℃，使互补氢键断裂，DNA 变性，两条链很快分开。如温度逐渐降低，则 DNA 复性，两条链的氢键又重新形成。假如有该 DNA 合成的 RNA 分子存在时，缓慢冷却就会出现DNA-RNA 杂交分子，就像 DNA 双螺旋复性那样。此一试验图示于 8-5，从试验中我们可以看到，只有一条 SP 8 的 DNA 链与 RNA 互补。

图 7-5 用 DNA-RNA 杂交链的检出试验说明，一个做模板的双链 DNA，与其合成的 RNA 之间，只有一条 DNA 链可与 RNA 链形成 DNA-RNA 杂交分子。图的左边说明，DNA 解链后混入其拷贝的 RNA，可形成 DNA-RNA 杂交物。右边的图说明，将 RNA 混入其非模板链的 DNA 中，无 DNA-RNA 杂交分子形成。

以上结果还解释了为什么双链 DNA 的碱基数目 A 等于 T，而 G 则等于 C，但产生了 RNA 后，这种比例就没有了。不过，单链 DNA 分子产生了 RNA 链时，A 与 U，G 与 C 则接近相等。

三、RNA 链按照一个固定的方向合成

5′端

3′端

RNA 聚合酶

(a)链生长的方向 5′→3′

焦磷酸酶

焦磷酸酶

RNA 聚合酶

(b)链生长的方向 3′→5′

RNA 聚合酶

(d)

3′-脱氧腺苷三磷酸

(c)3′-脱氧腺苷三磷酸掺入

图 7-6 RNA 合成的方向图

(d)无 3′-羟基团与下一个核苷三磷酸反应，链生长终止。

像每一条DNA链那样，每条RNA链均有被糖磷酸骨架所规定的走向。链的末端被5′碳原子结尾，称5′-末端；而含有3′碳原子的一端则称为3′-末端。就像DNA合成那样，RNA链的生长顺序只从两端中的一端开始是比较合理的，不是从5′-3′，就是从3′-5′。假如RNA生长是从5′→3′，则我们将预料核苷酸开始于一个Ⓟ～Ⓟ～Ⓟ(三磷酸)基团；假如链的生长是从3′→5′，则生长末端的核苷酸应含有Ⓟ～Ⓟ～Ⓟ基团。可靠的证据表明，生长的方向为5′→3′，新掺入的核苷酸都在3′-末端发现，而Ⓟ～Ⓟ～Ⓟ基团则在链开始生长的核苷酸上发现。

用代谢抑制剂，3′-脱氧腺苷可以证明RNA的合成方向是5′→3′。3′-脱氧腺苷与腺苷相似，但在3′处缺氧原子。当将3′-脱氧腺苷加入细胞中后，它首先被磷酸化成3′-脱氧腺苷-Ⓟ～Ⓟ～Ⓟ，然后连接到3′的生长端上去，因为它不含3′-OH基团，其它核苷三磷酸就不再能与其结合，RNA合成就终止。这就说明合成方向是5′→3′，如为3′→5′，抑制剂是不能掺入的。

图7-6(a)与(b)中说明，RNA合成的方向总是从5′→3′的。(c)说明，3′-脱氧腺苷三磷酸的结构及其掺入一条从5′→3′生长中的RNA链的3′-末端，及其使RNA合成停止的情况。

第四节　RNA聚合酶

一、RNA聚合酶的结构

像DNA聚合酶Ⅲ那样，细菌RNA聚合酶也有一个很复杂的亚基结构。在有活性的状态时，全酶的沉降系数约为15 S，含有5种不同的肽链，β'，β，σ，α和ω，它们的分子量顺次为：155，151，70，36.5和约为11 kD(以上均为大肠杆菌的)。各链之间没有共价键相连，仅靠次级键聚合。在有活性的分子中，其它亚基仅出现一次，但α例外，为两次。全酶为$\beta'\beta\alpha_2\omega\sigma$，分子量约为450 kD(见图7-7，表7-2)。假如此酶为球状，RNA聚合酶的直径大约为10 nm，有30 bp的双螺旋DNA链那样长。然而，实际上它与DNA结合时，覆盖60个核苷酸，这表明它具有延伸的形状。

人们企图使细菌RNA聚合酶的复杂性与RNA合成的起始、延伸和终止等各个酶促步骤联系起来。但噬菌体T 7编码的RNA聚合酶虽只含一条分子量为98 kD的链，却具有所有这些功能。细菌酶的结构为何如此复杂，想必是细菌的转录较为精确和进行调控所需要的。可惜至今尚不能深入了解细菌RNA聚合酶的三维结构，原因是此酶太大，难于进行X射线结晶学研究，也还没有制出此种酶的晶体。

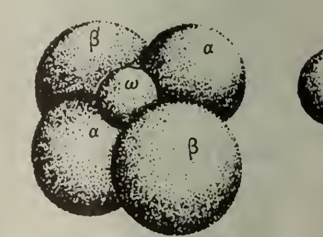


图7-7　RNA聚合酶的结构示意图

σ与其它亚基的结合不是特别牢固的，所以，它与$\beta'\beta\alpha\alpha\omega$团聚在一起的亚基很易分离。$\beta'\beta\alpha\alpha\omega$形成的聚集体称为核心酶(core enzyme)，因为不管是否有σ存在它同样有效地催化形成核苷酸之间的磷酸二酯键。催化中心看来在核心酶的β亚基上，β亚基也是抗菌素利福平(rifampicin)一类药剂的结合部位。利福平是此酶的有效抑制剂，它阻止RNA链的起始，但不影响以后的延伸。抗利福平的突变菌株，其突变位点即在编码β的基因*rpoB*位点上。β亚基基因的突变，是一种非致命的突变，对RNA聚合酶功能的影响很微妙。利链菌素(streptolydigin)也是一种抗菌素，它抑制RNA链的延伸，突变细菌对此抗菌素产生抗性，也是在编码β的基因上产生突变。

表 7-2 大肠杆菌 RNA 聚合酶的亚基和转录因子

基因名		基因图上的位置*(min)	多肽分子量(kD)	在酶中的数目	功 能
RNA 聚合酶亚基					
β'	*rpoC*	89.5	155	1	与 DNA 结合?
β	*rpoB*	89.5	151	1	RNA 聚合酶作用的催化部位
α	*rpoD*	66.5	70	1	识别启动子,并启动
α	*rpoA*	72	36.5	2	?
ω	—	—	11	1	?
转录因子					
ρ	*rho*	84.5	46	6	终止
nusA	*nusA*	65	69	1	延伸,终止

* 基因图指大肠杆菌的常规环形基因图。

二、RNA 聚合酶的基因

大肠杆菌 RNA 聚合酶亚基的基因,除了 ω 的基因未知之外,都在 3 个操纵子中找到,其中也含有其它一些基因(见图 7-8)。

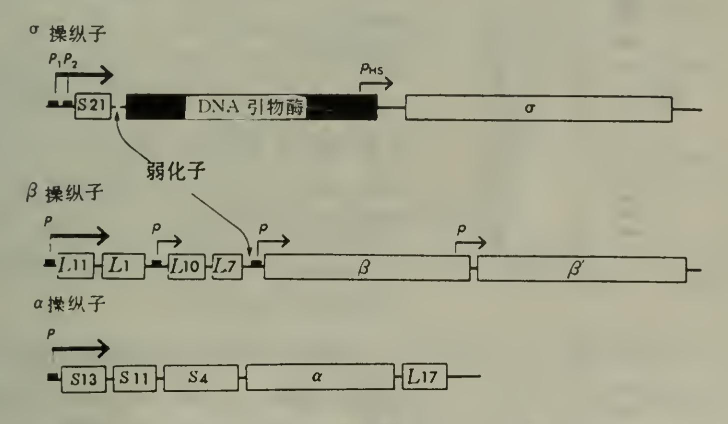


图 7-8 3 个含有 RNA 聚合酶亚基基因的操纵子图

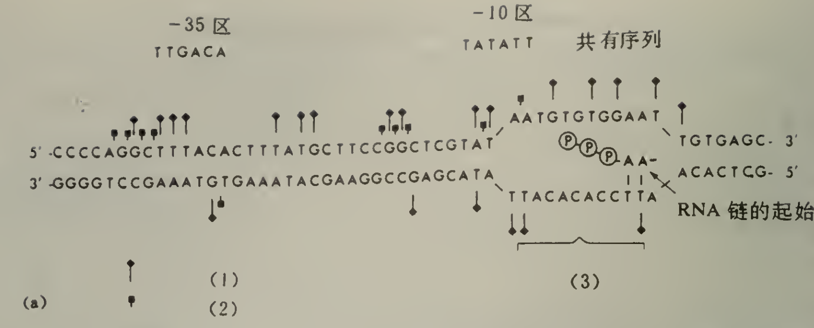
■示 DNA 聚合酶。

图 7-8 列出 3 个含有 RNA 聚合酶亚基 σ,β,β' 和 α 的基因的操纵子。主要的启动子(P)在左侧,RNA 向右合成。次要的启动子(P)可以在这些操纵子中找到,它们只能启动和转录它们右侧的基因,其中一个(p_{HS})含在 σ 操纵子中,它在热激条件下有活性。β 操纵子头两个基因 $L11$ 和 $L1$ 有时可以看成一个单独的操纵子,因为在基因 $L10$ 前有一个启动子。

这些基因与一些其它蛋白质基因组合在一起,很明显是共同调节的需要,这种需要与细胞的生长速度密切相关,但是这些操纵子是如何调节的尚不得而知。虽然知道 σ 操纵子有两个串连的启动子,就像核糖体 RNA 的操纵子一样,想必也是按生长速度受 ppGpp 分子控制的,ppGpp 是鸟嘌呤在其 5′端和 3′碳原子上各连有两个磷酸,过去也写成Ⓟ～Ⓟ—G～Ⓟ～Ⓟ。这些基因在这些操纵子中,都不是平均表达的。

三、RNA 聚合酶的识别功能与启动子

σ 亚基本身并无催化功能,其作用为识别DNA分子上的起始信号。体外试验证明,缺乏 σ 的核心酶只会偶然性地启动 RNA 的合成,但这是错误启动的结果。当 RNA 链只被核心酶



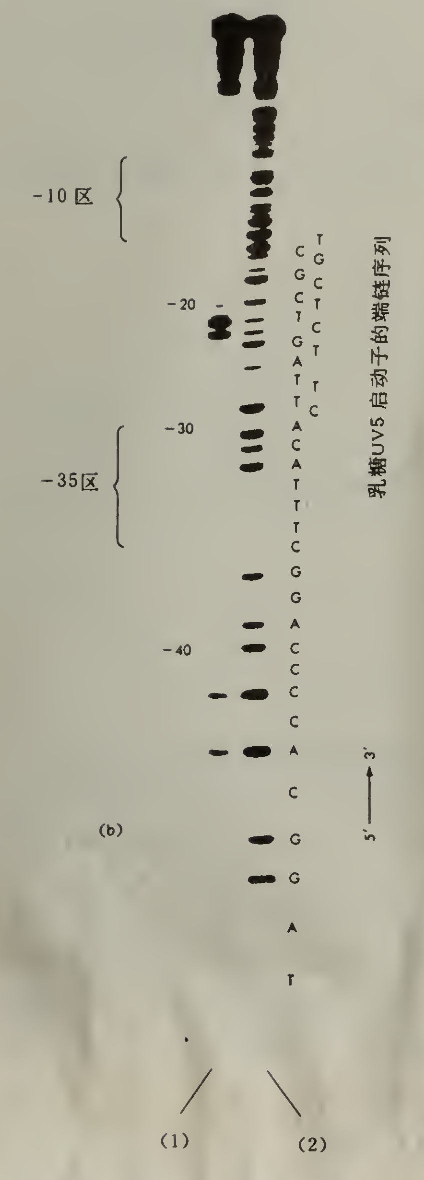


图 7-9 启动子的共有序列和 RNA 聚合酶的结合部位

合成时，则会随机地在一个基因的两条链上启动。当有 σ 时，就会选择正确的位点，全酶能够专一地与一个"启动序列"结合，这个序列就是启动子。

DNA 结构的何种特征引导细菌 RNA 聚合酶仅结合于启动子上呢？显然，某些核苷酸序列使 DNA 产生能与 RNA 聚合酶契合的局部构象，正如酶与自己的底物相结合一样。对 100 多个启动子做过的比较说明，大约在 RNA 合成起点之前的 10 bp 和 35 bp 处，有两个由 6 个核苷酸组成的共有序列，这些碱基被命名为－35 和－10 序列，这是将编码 RNA 链的头一个 DNA 核苷酸定为＋1 来编号的。少数启动子正好具有－10 区和－35 区的共有序列(consensus sequence)，大多数共有序列仅有少数几个核苷酸不同(图 7-9)。

图 7-9 表示启动子的共有序列和 RNA 聚合酶的接触位点。(a)在－35 区和－10 区共有序列的下面是一个称为乳糖 UV5 的启动子。RNA 聚合酶与启动子中，尤其－35 和－10 附近的碱基[上画有标记符号和分子骨架中的磷酸基(符号标在碱基中间)]相接触。(1)被 RNA 聚合酶保护，免于受化学或光化学作用的碱基；(2)在开放复合物中，与 RNA 聚合酶结合的磷酸；(3)开放复合物中，被 RNA 聚合酶解链的区域。本次试验表明至少有 12 bp 的片段解链，有的试验指出解链片段为 17 个碱基。(b)乳糖 UV5 启动子的"足迹分析"(footprint analysis)。(1)在用限制性脱氧核糖核酸酶处理前启动子先与 RNA 聚合酶结合；(2)在没有 RNA 聚合酶的条件下用脱氧核糖核酸酶处理。受 RNA 聚合酶保护的 DNA 序列能免受 DNA 酶的限制性核酸酶消化。这个实验说明，RNA 聚合酶覆盖了 50～60 个 DNA 碱基，覆盖部位使核酸酶不能靠近和消化磷酸二酯键。

“保护”试验,“化学修饰”试验等,强有力地证明,这些序列是启动子的关键部分。第一,RNA 聚合酶是在－35 区和－10 区序列中与磷酸与碱基接触;第二,较弱的启动子与共有序列区别较大,较强启动子与共有序列更为接近;第三,此点最明显,即 75%的能摧毁启动子功能的突变都发生在共有序列中,其余 25%的变化也在共有序列附近发生(见图 7-10)。

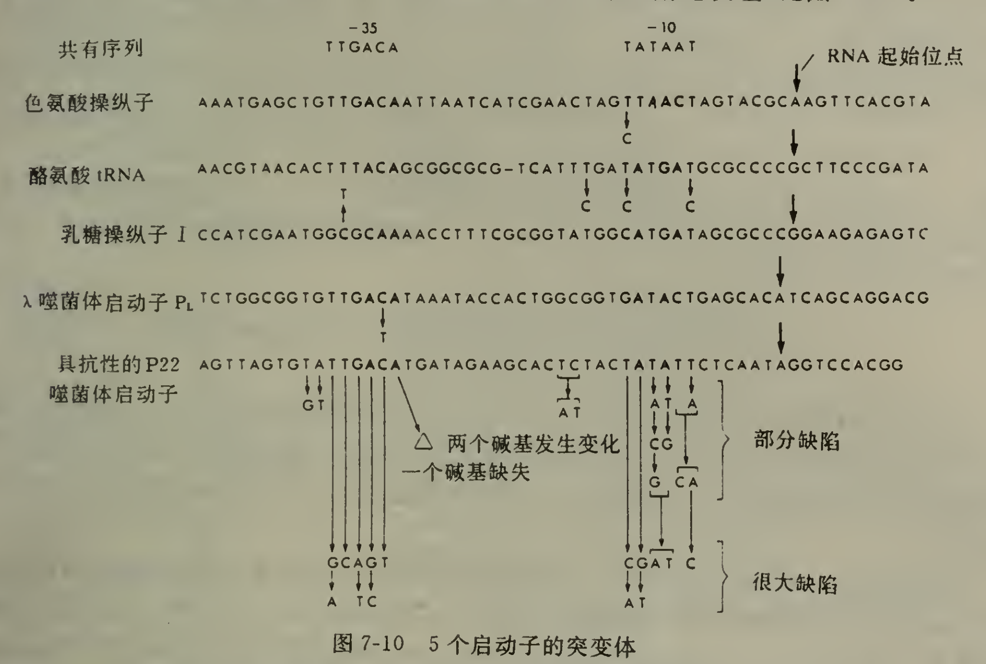


图 7-10　5 个启动子的突变体

图 7-10 是 5 个启动子的突变体。突变的减少,用向下的箭头表示;突变的增加,用向上的箭头表示。

两个结合区距离大约 20 bp,或者是双螺旋两个整圈。由于两个结合区均在 DNA 分子的同侧,酶必须与螺旋的同一侧面结合,酶还必须识别每一结合区的专一结构,这些专一的结构都是在凹槽的底部由碱基对构成的(见图 7-11)。

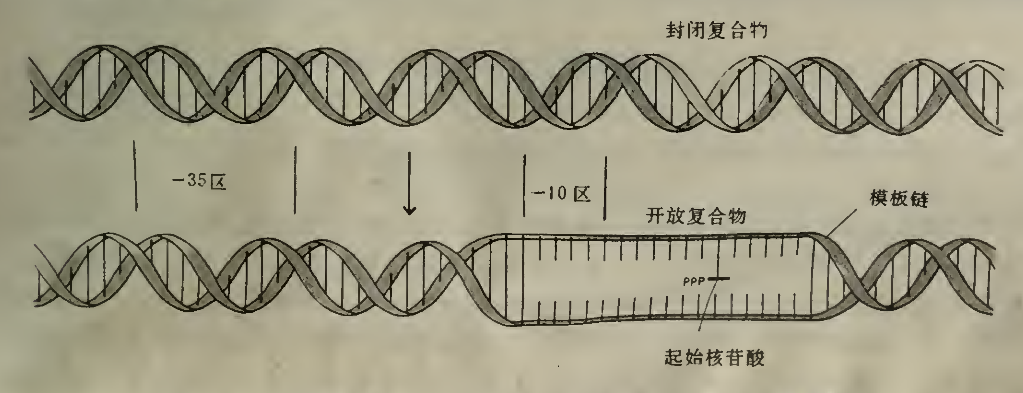


图 7-11　RNA 聚合酶与 DNA 结合形成封闭和开放的启动子复合物

图 7-11 说明,RNA 聚合酶与启动子刚开始结合时是封闭复合物;当 DNA 链分开,RNA 链开始延伸就形成了开放复合物。RNA 链的起始三磷酸核苷也在图中绘出。

四、RNA 聚合酶与 DNA 结合并使之解链

大肠杆菌RNA 聚合酶与一个启动子结合有两个明显的步骤(见图 7-11)。首先,酶“发现”启动子的序列并与之结合成较松弛的“封闭型”复合体。此一步骤大都发生在－35 区,改变了－35 区碱基序列的突变体,就降低了这种结合功能;第二,封闭型复合体转变成“开放型”复合体,此时 RNA 聚合酶与 DNA 结合得更紧密。开放型复合体在－10 区解链,解链范围约为 17 bp,解链有使模板暴露的作用,在暴露的模板上聚集了等待合成 RNA 的核苷酸。通过研究发现,－10 区内含有丰富的 AT 碱基对,由于 AT 融解比 GC 容易,故对解链有利;进一步研究证明,－10 区内的碱基序列常有改变,但 AT 含量保持不变。－10 区内碱基序列突变,则开放型复合物不能形成。开放型复合物一旦形成,DNA 就继续解链,将解链区向启动子－10 区以外远离启动子的方向移动。

五、不同的 σ 因子使 RNA 聚合酶识别不同的启动子序列

RNA 聚合酶的核心酶能够催化 RNA 的合成,但要识别启动子序列,却一定要有 σ 因子。这一结论的证据来自于不同 σ 因子的发现。这些不同的起主要作用的 σ 因子可以指导 RNA 聚合酶作用于不同的启动子上,这些非常 σ 因子识别的启动子序列和主要的启动子序列有很大的区别。

这类特殊的 σ 因子在细胞中执行一些特殊功能,使某一些或某一套基因能够表达。首先被发现的是枯草芽胞杆菌中负责生成芽胞的 σ 因子。在大肠杆菌中,有一种 σ 因子(σ^{32}),能够识别热激蛋白基因的启动子,它启动热激基因的表达,当细胞周围的温度急剧上升,热激基因表达后就产生热激蛋白质。还有一些噬菌体产生的 σ 因子,在噬菌体发育时,为噬菌体转录产生必需的蛋白质。

最近的研究说明,噬菌体内编码合成的 σ 因子,在指导合成大肠杆菌噬菌体 T4 晚期基因时,σ 因子只与－10 区结合。T4 晚期启动子的－35 区序列可以用基因工程的方法改变而不会损害启动子的功能。这说明 σ 对－35 区的结合可能是不需要的,需要结合的是－10 区。

第五节 启 动 子

一、启动子的不同效率

从上述资料可以看出,各启动子的效率是不一样的,在一些序列上转录得好一些;另一些较差。转录得好的启动子可以使RNA 合成从开始到终止重复得更快一些(即轮换速度快)。生化试验说明,启动子开始与酶结合和开放型复合物的形成这两个步骤,在启动子之间效率是不相同的。因此,细胞中的启动子在启动速度上可以划分成不同的等级,有的是 10 min 或 10 多分钟启动一次,有的仅 1～2 s 内启动一次。这是一个基本的限速步骤,在这一步骤的基础上,基因表达的速度就确定了。

二、调节蛋白对启动子的影响

一个特定基因的转录速度不是固定不变的,它需要在不同的生长环境中根据情况来改变

转录的节奏，这就是基因的调节。在大肠杆菌中，转录的调节经常有蛋白质介入，这种蛋白质在启动子附近，或在启动子内部，它们起增加或降低 RNA 聚合酶启动 RNA 合成速度的作用。

这些蛋白质包括阻遏物，它阻碍 RNA 聚合酶的结合，从而抑制转录；还包括激活剂，它与 RNA 聚合酶结合并刺激其活性。需要激活剂的启动子常与－35 区的共有序列结合较弱，看来激活剂是补足结合位点的这一功能。还有一种情况是，一个蛋白质是一个启动子的阻遏物，也可能是另一个启动子的激活剂，也有反过来的情况。根据调节蛋白首次发现的位置和功能，往往以该情况下的条件命名。某些重要的调节蛋白载于表 7-3 中。

表 7-3　影响启动子功能的调节蛋白

蛋白质	对靶启动子起始的效果	在细胞中的功能
乳糖阻遏物	阻止	细胞缺糖时，阻止合成代谢乳糖的酶
半乳糖阻遏物	阻止	细胞缺糖时，阻止合成代谢半乳糖的酶
LexA 阻遏物	阻止	细胞 DNA 无损伤时，阻止合成 DNA 修复酶
λ 阻遏物	阻止 (P_L 与 P_R 启动子)	阻止合成为溶菌生长所需的蛋白质
λ 阻遏物	激活 (P_{RM}启动子)	用刺激本身启动子的办法来增加自身的合成
CAP 蛋白	阻止	葡萄糖缺乏时，调节自身的基因，并且阻止其它不需要蛋白质的合成
CAP 蛋白	激活	葡萄糖不能利用时，增加代谢其它糖的酶的合成
AraC 蛋白	阻止	细胞无糖时，阻止代谢阿拉伯糖的酶的合成
AraC 蛋白	激活	有阿拉伯糖可利用时，增加利用阿拉伯糖所需的蛋白质的合成
λcⅡ 蛋白	激活	刺激建立溶源性所需要的噬菌体蛋白质的合成

三、上游 DNA 序列可增强启动子的活性

虽然启动子的－10 区和－35 区中的 6 个碱基序列是 RNA 聚合酶结合和启动的必需序列，但它们附近的其它 DNA 序列也能影响启动子的功能。例如，在启动子位点上游 50～150 核苷酸之间的自然序列在合成核糖体 RNA 时能增强启动子的活性。假如这些序列缺失，或以外源 DNA 代替（如一个质粒 DNA 进入此一序列中），则此核糖体 RNA 基因在进行克隆和表达时，启动的速度将降低 10 倍。

另外，上游序列在某些时候可以作为一些激活剂的结合部位，直接激活 RNA 聚合酶。有时启动子上游序列还可以和一个拓扑异构酶结合，此酶可以诱导 DNA 形成有利的超螺旋状态，以利 RNA 的合成。上游序列还可帮助 DNA，使 RNA 聚合酶能更好地接近 DNA 的启动区段，或者使 DNA 与蛋白质结合固定在细胞结构上。最后存在着这样的可能，一些专一性序列所引起的 DNA 结构上的微妙差别会沿螺旋传递很长距离，所以，某一特定的上游序列可以影响－10 和－35 区的 DNA 结构。

四、模板的超螺旋张力对启动子效率的影响

由于 RNA 聚合酶在开始合成 RNA 之前，必须在启动子处解开 DNA 链，有助于或妨碍解链的条件必然增加或降低 RNA 合成的起始速度。在细胞中 DNA 模板超螺旋的张力是影响起始的重要条件之一，DNA 旋转酶（gyrase）使 DNA 产生负超螺旋作用，此作用有利于解链。但负超螺旋作用又受拓扑异构酶 I（topoisomerase I）的抗衡，它使过度的负超螺旋松弛。上述两种酶的调节，可使 DNA 达到解链的最适状态。突变或抑制剂引起的上述酶活性的改变都能

使启动子的功能向所希望的方向发展。如旋转酶抑制剂萘啶酮酸(nalidixic acid)可抑制许多基因的表达。使编码拓扑异构酶I的基因突变,可产生相反的效果,增加了某些基因的表达。当细胞中的DNA的负超螺旋充分时,DNA旋转酶的合成就降低了,这是细胞的自我调节。

第六节 RNA的酶促合成

一、RNA合成的起始

(一)RNA链的合成从pppA或pppG开始

RNA合成的开始,实际上发生在解链区段的一端,此端大肠杆菌RNA聚合酶与之结合,结合位置在解链区段之内(图7-9与图7-11),大约在自-10区起第12或13个碱基处。第一个核苷酸常常是pppA或pppG,有时是pppC,pppU很少见。有时酶一开始偶然落在附近几个核苷酸上,这时酶就扫描解链区,寻找一个嘌呤做起点,如无嘌呤,第二位则选择嘧啶做起点。

从细胞中分离出来的RNA分子,其5′-末端的核苷酸不是开始转录时的核苷酸,因为在RNA合成之后,要被核酸内切酶切去一小段,剩下的链端很多是从嘧啶开始的。例如,tRNA链从U开始,它们的末端也无三磷酸基团,这是长链在体内切开形成的。实际上,所有细胞都含有各种核酸酶,有关起始核苷酸序列的资料大都来自体外试验。

(二)在第一个核苷酸参加合成之后σ就解离

RNA链开始延长后,σ从核心酶-DNA-新生RNA链构成的复合物上解离下来,随时能附着在另外的核心酶分子上(见图7-12)。

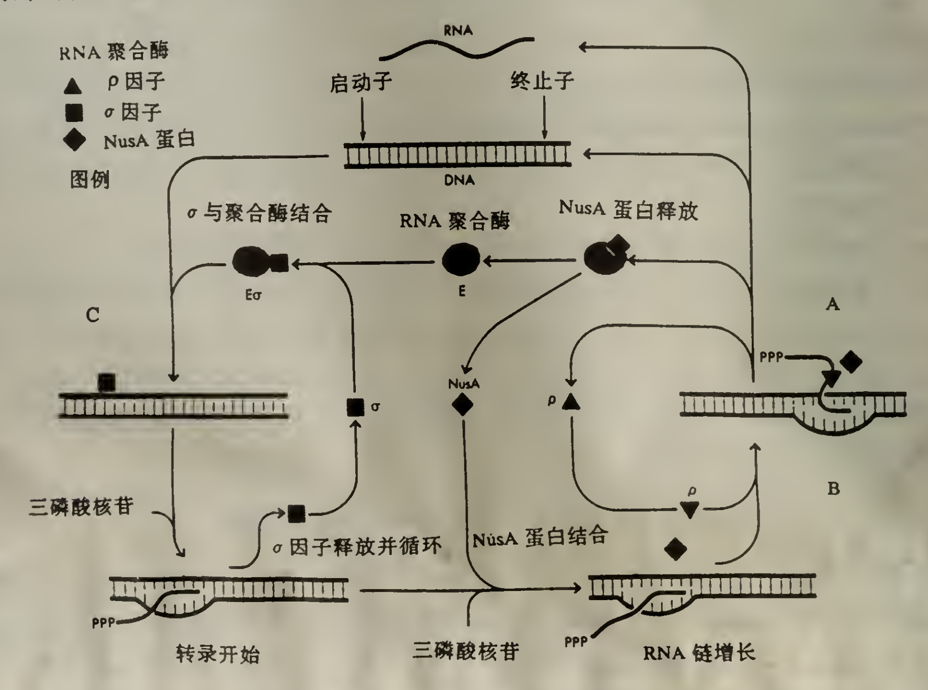


图7-12 RNA的合成示意图

A. ρ因子,RNA和RNA聚合酶均释放;B. ρ因子与RNA和RNA聚合酶结合;C. σ帮助识别转录的起始点。

在图 7-12 的情况下，链的终止需要 ρ 因子，在另外的情况下，RNA 聚合酶自己能阅读终止信号。

可以想象，σ 的释放可能是它的功能不再需要了，或者是 σ 的继续存在可能会因与启动子序列紧密结合从而使酶（核心酶）以后难于沿着模板滑动。

全酶与无启动子序列的 DNA 结合时，σ 因子抑制了 RNA 聚合酶与 DNA 的非专一性结合，它使酶轻轻沿着 DNA 滑动，直至找到一个启动子序列，σ 因子就使全酶与启动子紧密地结合。σ 释放后，核心酶与启动子的结合就减弱。

由于正在转录的 RNA 聚合酶分子缺乏 σ 和可能只有一半的酶分子在细胞中是活跃的，因此，σ 因子与酶分子的数量不需要相等，不过它们也不会相差太悬殊，也可能游离的 σ 保持为一个库，所以，当核心酶在终端位置上释放出来后，又很快重新形成了全酶。

二、RNA 合成链的延伸

（一）当酶前进时 DNA 被 RNA 聚合酶解链然后又复链

当酶使 RNA 链延伸时，RNA 聚合酶继续解开 DNA 链，使模板暴露出来。RNA 链开始的部分是一个 RNA-DNA 杂交区，其长度约为 12 个碱基，当 RNA 链离开模板链后，两条 DNA 又重新复链成为双螺旋（见图 7-13）。

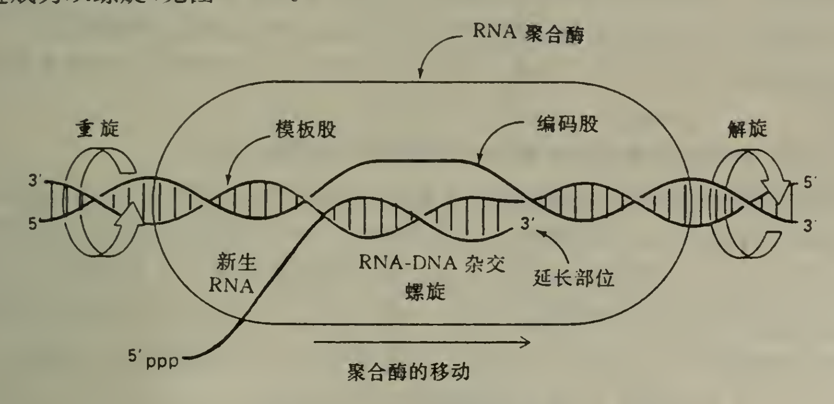


图 7-13 RNA 延伸的图解

DNA 的解链部分要比 RNA-DNA 杂交物的长度长，约为 17 个碱基，与开放复合物在启动子处形成时的长度相同。DNA 被 RNA 聚合酶解链的长度测量起来很方便，方法是用一个拓扑异构酶去松弛正在被转录的共价闭环 DNA，然后加入去污剂去掉 RNA 聚合酶和 RNA，最后测量 DNA 中超螺旋的数目即可（见图 7-14）。拓扑异构酶作用时，每解开 10.4 个碱基就会出现一个新的负超螺旋。于是，根据活性酶的数目我们就能计算出每一次解链的碱基对数目。

在图 7-14 中，(a)说明，当 DNA 转录时，用拓扑异构酶去松弛共价闭环 DNA，人们就能捕获活性 RNA 聚合酶分子所诱导的解链程度。使 RNA 聚合酶变性，转录复合物瓦解，部分解链 DNA 变成负超螺旋，变得更紧密，在琼脂糖凝胶中负超螺旋的迁移比松弛的 DNA 快；(b)将诱导出来的超螺旋数目与活性的 RNA 聚合酶的分子数目做图，可以看出每个酶解链的数量是：每个酶为 1.6 个超螺旋，或每个酶解链 17 bp（DNA 1.6×10.5 bp）。

虽然延伸复合物一般使 DNA 解开 17 bp，但当 RNA 聚合酶遇到特殊的 DNA 序列时，这

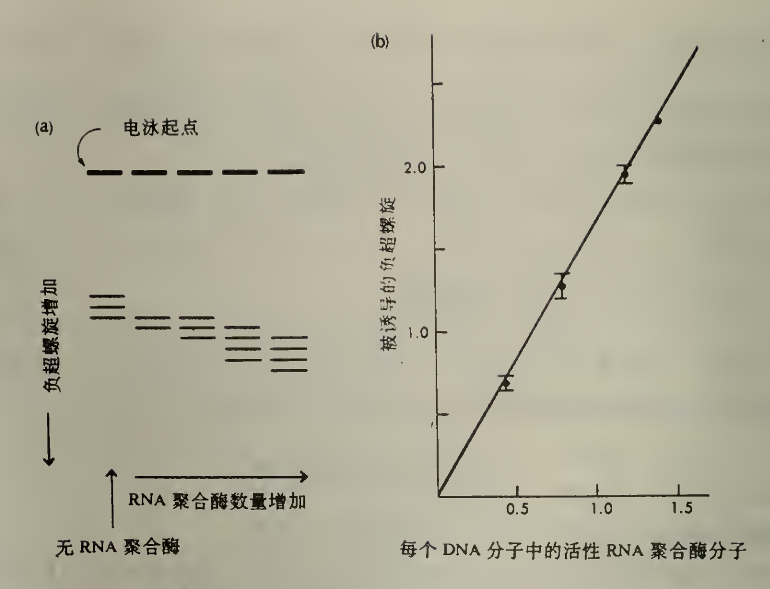


图 7-14 被 RNA 聚合酶诱导解链的 DNA

个数目变化就很大。例如，在终止处时，酶和 RNA 链从复合物上释放下来，DNA 就复链。相反的情况发生在转录 *Col*EI 质粒复制所需的 RNA 引物中，DNA 链的重新盘绕在专一位点受到阻碍，转录本就保持为 RNA-DNA 杂交分子，它可长达几百个核苷酸，此种现象是如何发生的尚不得而知。

（二）在 RNA 合成的延伸和终止时 NusA 因子可以代替 σ 因子

一个 RNA 聚合酶附属因子，称为 NusA 蛋白，经很多年它才被认识。因为此因子对酶的基本活力无影响，纯化 RNA 聚合酶时又未发现酶与 NusA 蛋白在一起，因此一直对它忽视。但此因子对 RNA 的合成很重要，它是在 NusA 蛋白紧紧地结合在 λ 基因蛋白上时发现的，后者是 λ 生长中防止转录终止的调节因子。像 σ 一样，NusA 与 RNA 聚合酶的核心酶结合，但不那么牢固，所以在纯化时，σ 就取代了 NusA。这些特性说明，在细胞中当 σ 释放出来后，NusA 就马上与这个 RNA 聚合酶分子结合，并制造出一条 RNA 链。但是，在转录时，NusA 并非总是固定在一个 RNA 聚合酶上，帮助延伸一条 RNA 链，在转录中它可以和几个 RNA 聚合酶分子起作用。因此，细胞中的 NusA 分子并不比纯化时所获得的多多少，每个细胞才几百个，但却能与所有执行转录功能的 RNA 聚合酶分子起作用。可以预料，RNA 聚合酶从 DNA 上释放出来后，σ 因子由于与核心酶的亲和力比 NusA 大，所以 σ 又取代了 NusA 与核心酶结合在一起（图 7-12）。

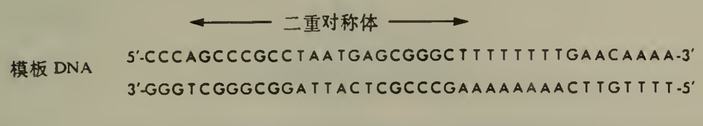
NusA 蛋白在转录中的全部作用目前很难说清楚，虽然遗传试验说明它是一种重要的蛋白质，但具体的功能尚属未知。在体外，NusA 对 RNA 聚合酶在转录的速度和终止方面都显示出肯定的作用，这些肯定的作用大概是在调节因子之间起中介的机制，就像在 λ 噬菌体 *N* 基因蛋白上的作用那样。

三、RNA 合成的终止

在细菌 DNA 中，有一个信号称为终止子（terminators），其功能为在 DNA 模板上的一个特定点停止 RNA 的合成。在终止点上，RNA 聚合酶停止使核苷酸聚合成 RNA，释放 RNA

链，使DNA 在另外的启动子上重新开始转录。在某些位点上，终止需要一个附属蛋白，即 ρ 因子，在另外一些位点上，核心酶本身即能实现 RNA 合成的终止。

对于细菌及病毒 DNA 的终止子序列做过很多分析和研究，特别是对于没有 ρ 因子的序列着重进行了研究。在不依靠 ρ 因子的终止方式上发现了两个明显的特点：即在 DNA 上有一个二重对称性(dyad symmetry)，此对称性集中在 15～20 个核苷酸上，位于 RNA 尾部之前，另一特点为在 DNA 模板链上有一串大约为 6 个 A 的碱基，它们转录为 RNA 末端的一连串的 U(见图 7-15)。



转录本(RNA) 5'-CCCAGCCCGCCUAAUGAGCGGGCUUUUUUUU-OH 3'

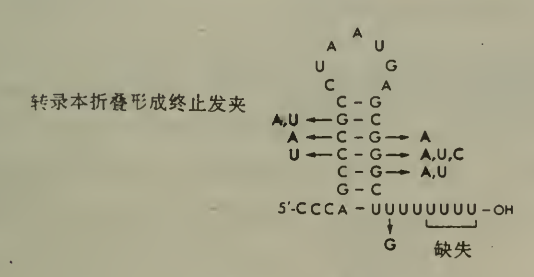


图 7-15　不依赖 ρ 因子的终止序列和终止 RNA 的结构

图中箭头处为突变碱基，缺失碱基已标明。

虽然图中的二重对称性往往是 DNA 与寡聚蛋白结合的标志，但转录本形成的发夹结构对终止来说，可能更为重要(见图 7-16)。

体外试验说明，终止可由掺入不能按正常方式配对的改变碱基而防止，这就证明发夹结构的形成对终止是重要的。此外，大多数丧失终止功能的突变体涉及单一核苷酸的改变，从而破坏发夹结构中双螺旋茎的形成(图 7-15)。终止子突变体的分析也指出了模板链多聚 dA 的重要性，这一序列中某一碱基的改变或缺失部分碱基都会使终止子失活。

图 7-16 表示，RNA 的一段短的双螺旋和富含 U 的序列如何导致终止作用和 RNA 链的释放的。(a)是延伸复合物恰好完成富含 U 的 RNA 链；(b)RNA-RNA 杂交物(发夹)的形成破坏了一部分 RNA-DNA 杂交链，仅留下多聚 U 与多聚 A 的一段杂交链；(c)多聚 U 与多聚 A 杂交物解离，转录本即被释放。

图 7-16 的模型说明，发夹结构的形成，对于终止子的作用，其重要性是很清楚的。

依赖 ρ 因子的终止子，没有不依赖 ρ 因子终止子那样的多聚 A 序列特征，而且也不一定形成稳定的发夹。至今尚不知道依赖 ρ 因子的终止子是如何以 ρ 因子来代替这些发夹结构序列的作用的。但存在下面一些可供参考的线索：第一，ρ 在 RNA 存在下对 ATP 的水解提示，它可能结合在新生的 RNA 链上并利用 ATP 的能量，使 ρ 产生将 RNA 链从酶和模板链中释放出来的活性。第二，ρ 可能是 RNA 聚合酶在起作用时与该酶结合，因为某些因 *rho* 基因突变而发生改变的 ρ 蛋白，仅仅在编码 RNA 聚合酶 β 亚基基因发生专一性突变从而受到修饰后才具有终止活性。第三，我们知道，RNA 聚合酶本身能识别依赖 ρ 的 DNA 终止序列，这以后 ρ 就释放出 RNA。证据是 RNA 聚合酶本身在需要 ρ 的终止处停顿，如反应中无 ρ 存在则 RNA

聚合酶会一直移动下去直到 ρ 的出现。一种阐明 ρ 功能的模式是，即便是一个弱发夹结构也会使 RNA 聚合酶的移动停顿，使 ρ 得以结合和使 RNA 与聚合酶实现酶促分离。

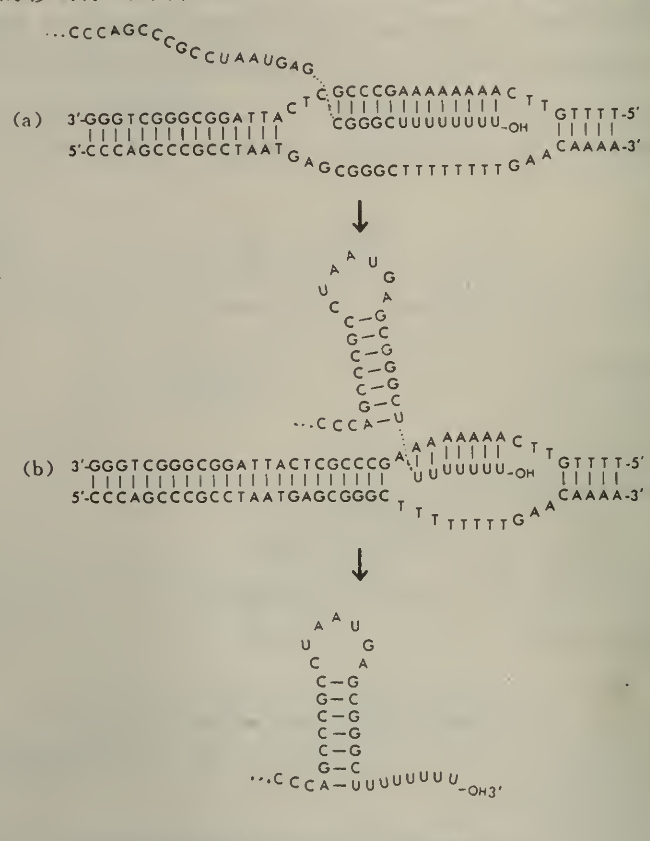


图 7-16　RNA 合成时形成的发夹终止结构

第七节　RNA 转录后的加工

RNA 在转录完成后，继续形成具有功能的活性 RNA 分子。在原核细胞中，加工方式可以分为以下 3 种类型。

一、tRNA 及 rRNA 分子由某些新生 RNA 链的裂解及化学修饰形成

大肠杆菌的 3 种 rRNA 及一种 tRNA 分子是由含有间隔区的初级 RNA 转录本剪切形成（见图 7-17）。

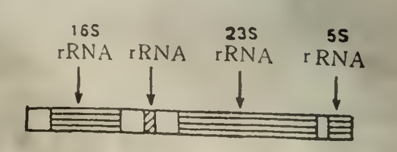


图 7-17　rRNA 的初级转录本

初级转录本的裂解产生 5 S、16 S 及 23 S rRNA 分子及一个 tRNA 分子。空白框为间隔区。

另外，一些转录本含有几种 tRNA 的序列，或同一 tRNA 的几个拷贝。裂解及剪切这些 rRNA 及tRNA前体的核酸酶是高度精确的。例如，大肠杆菌的核糖核酸酶 P 能正确地制成其全部 tRNA 分子的 5′-末端。核酸

酶Ⅲ通过裂解初级转录本发夹区的专一位点而切下5 S,16 S及23 S rRNA前体。相反的是,原核细胞的mRNA分子不经过或很少被修饰。事实上是这些mRNA在转录时,有的就同时被翻译。但是,有些病毒的mRNA(如T7噬菌体mRNA),在翻译前也要被核酸酶Ⅲ裂解。

二、在有些RNA链末端加上核苷酸

很多情况是将CCA加到原来没有此序列的tRNA分子的3′-末端上。在真核细胞中则是将长段的多聚A加在大多数mRNA分子的3′-末端上,将甲基化的鸟核苷酸加在5′-末端上(称为帽子,以后将要叙述)。

三、碱基和核糖的修饰

在真核细胞的rRNA中,100个核糖单位大约有一个核糖的2′-OH由S-腺苷甲硫氨酸经酶促被甲基化。在细菌中则是rRNA的碱基而非核糖被甲基化。还有一个普遍现象是所有的tRNA分子中均含有稀有碱基,它们是由tRNA前体中的常见核苷酸经酶促修饰而成。例如,假尿苷酸及胸腺嘧啶核苷是尿嘧啶核苷酸残基在转录后修饰而成。

在真核细胞中,所有RNA转录本皆需普遍加工。rRNA和tRNA的裂解与修饰类似原核细胞。与原核细胞最不同的是,真核细胞的mRNA来自一个很大的转录本,经裂解与拼接而成。真核细胞的转录与翻译在细胞中的不同区域进行。似乎RNA的加工对调节mRNA,rRNA和tRNA从胞核到胞浆的转运上起了关键作用。

第八节　真核生物RNA的合成

一、真核生物的RNA聚合酶

真核细胞中有3种RNA聚合酶,即RNA聚合酶Ⅰ,RNA聚合酶Ⅱ,RNA聚合酶Ⅲ。这些名称最早是依据它们从DEAE-纤维素柱上洗脱的先后顺序而定出来的。后来发现不同的生物3种RNA聚合酶的洗脱顺序不一样,后来就改为以3种酶对α-鹅膏蕈碱(α-amanitine)的敏感性不同来区别。

RNA聚合酶Ⅰ:基本不受α-鹅膏蕈碱的抑制,在大于10^{-3}mol/L时才表现出轻微的抑制作用。此酶存在于核仁中,其功能是合成5.8 S rRNA,18 S rRNA和28 S rRNA。

RNA聚合酶Ⅱ:对α-鹅膏蕈碱最为敏感,在10^{-9}～10^{-8}mol/L浓度下就会被抑制。此酶存在于核质中,其功能是合成mRNA和snRNA。

RNA聚合酶Ⅲ:对α-鹅膏蕈碱的敏感性介于酶Ⅰ及酶Ⅱ之间,在10^{-5}～10^{-4}mol/L时表现出抑制作用。此酶也存在于核质中,其功能是合成tRNA和5 S rRNA及转录Alu序列。在细胞质中也能发现一些RNA聚合酶Ⅲ,是从细胞核中渗漏出来的。

以上3种RNA聚合酶的分子量都在500 kD左右(14～15 S),每种酶分子含有两个大亚基和4～8个小亚基,每个小亚基的分子量为10～90 kD,各个亚基的功能还没有很好地了解。真核生物也像原核生物一样,不同种类的基因,需要不同的蛋白辅助因子协助RNA聚合酶进行工作。在真核生物的线粒体和叶绿体中,均已发现少数RNA聚合酶,它们分子量小,活性也较低。这些RNA聚合酶都是核基因编码,并在细胞质中合成后运送到细胞器中去的。

二、真核生物的启动子

真核生物有 3 种 RNA 聚合酶，每种酶都有自己的启动子。

（一）RNA 聚合酶Ⅰ的启动子

RNA 聚合酶Ⅰ是转录 rRNA 的，其启动子可分两部分：－40～＋5 称近启动子，其功能为决定转录起始的精确位置。－165～－40 称为远启动子，其功能为影响转录的频率。每种生物都有特定的转录因子与 RNA 聚合酶Ⅰ结合，从而促进酶与启动子结合形成转录起始复合物，因此，RNA 聚合酶Ⅰ具有明显的种族特异性。启动子均位于不转录的间隔区（nontranscribed spacer，NTS）。应该指出，不转录间隔区目前发现大部分是转录的，故系误称，但已沿用。

（二）RNA 聚合酶Ⅱ的启动子

真核生物 RNA 聚合酶Ⅱ的启动子是多部位结构，主要有四个部位，其中前三个部位是大多数启动子都具备的。

第一个部位为帽子位点，即转录起始位点。其碱基大多为 A（指非模板链），两侧各有若干个嘧啶核苷酸。

第二个部位为 TATA 框（Hogness box）。其共有序列为：$T82A97T93A85^{A63}_{T37}A83^{A50}_{T37}$，基本上由 AT 碱基对所组成，只有很少启动子中有 GC 碱基对的存在。

第三个部位为 CAAT 框。其共有序列为 $GG^{C}_{T}CAATCT$，一般位于－75 附近。虽然此序列名为 CAAT 框，但其中头两个 G 的重要性并不亚于 CAAT 部分。

第四个部位是增强子（enhancer），又称远上游序列（far upstream sequence）。一般都在－100以上。增强子的作用主要是，对依赖于 TATA 框的转录和不依赖 TATA 框的转录都有增强效应，但对前者增强效应高；距离效应，离 72 bp 的重复序列近的容易起始转录；极性效应，转录方向离开 72 bp 重复序列的起始序列优先转录，而朝向 72 bp 重复序列的则效率差；细胞类型的选择，不同细胞类型增强作用不一样。

（三）RNA 聚合酶Ⅲ的下游启动子

RNA 聚合酶Ⅲ的启动子在转录起始位点的下游，是在研究非洲爪蟾 5 S RNA 基因时发现的，故又称内部启动子。非洲爪蟾 5 S RNA 基因的启动子在转录起始点 50 bp 之后。在研究中发现，缺失＋50 以前的序列和缺失＋83 以后的序列都能正常转录，而唯独＋50～＋83 这段序列不可缺少。把这段序列插入任何 DNA 中，RNA 聚合酶Ⅲ都能识别并起始转录，转录起始点在其上游方向 50 bp 左右一个嘌呤碱基。这段序列就是5 S RNA的启动子。

除了 5 S RNA 基因外，RNA 聚合酶Ⅲ也能转录 tRNA 基因，部分 snRNA 基因和腺病毒的 *VA* 基因。这些基因也是内部启动子，但这些基因与 5 S RNA 启动子不同，是由两个不连续的区域所组成，靠近 5′方向的称为 A 区，靠近 3′方向的称为 B 区。这些启动子的结构见图 13-66。

RNA 聚合酶Ⅲ识别不同类别的启动子，在于有共同的转录因子 TFⅢC 和 TFⅢB，对于 tRNA 基因和 *VA*RNA 基因来说，只需要前述两种转录因子，而对于 5 S RNA 基因的启动子，除了这两种转录因子外，还需要一种 37 kD 专一的转录因子 TFⅢA，这因子与 RNA 聚合酶Ⅲ和启动子相互作用。

三、真核生物转录的起始和延伸

关于真核生物转录的起始机制，现在还不清楚，其原因在于转录机制十分复杂，以 RNA

聚合酶Ⅱ的转录起始尤甚。从启动子结构方面来说，它包含了许多启动子成分，每种成分均有相应的蛋白因子与之结合，这些蛋白因子总称为转录因子，目前均在分门别类的研究之中。

RNA 合成的速度大约为每秒 30～50 个核苷酸，但链的延长并非以恒定速度进行，有时会降低速度或延迟，这是延长阶段的重要特点，其原因尚不明。人们发现在通过一个富含 G·C 对的模板以后约 8～10 个碱基，则会出现一次延迟。如果在突变体中 G·C→A·T 则减少延迟。如果在连续的 G·C 对之间只有一个 A·T 对，把这个 A·T 变为 G·C 则会出现强烈的延迟作用。这种延迟作用可能对 RNA 链的终止和释放有关。

四、真核生物转录的终止

对于真核生物转录的终止信号和机制了解很少，其主要困难在于很难确定原初转录本的 3′-末端，因为大多数在转录后很快进行加工，无论是 mRNA，tRNA 还是 rRNA 都如此。病毒 SV40 的终止点经过研究，很像大肠杆菌不依赖 ρ 因子的终止子，有一个发夹结构，末端有一长序列 U。

对于 RNA 聚合酶Ⅲ，体外与体内转录本相同，表明体内转录确是在 RNA 末端处终止的。爪蟾 5 S RNA 的 3′-末端为 4 个 U，而这 4 个 U 的前后均为富含 G·C 的序列，在 G·C 序列中分布着寡聚 T(4 个以上)是所有真核生物 RNA 聚合酶Ⅲ转录的终止信号。这种序列特征高度保守，从酵母到人都很相似。

五、真核生物转录后的加工

(一)mRNA 的转录后加工

mRNA 转录后的加工主要是帽子及多聚(A)尾。N^7-甲基鸟嘌呤核苷酸部分称为帽子 O，符号为 m^7GpppX，单细胞真核生物如酵母，只具有帽子 O。如果在原来转录本的第一个核苷酸的 2′-O 位上产生甲基化，则构成帽子 1(1 个甲基化，包括帽子 O 部分)，其符号为 $m^7GpppXm$，这是除了单细胞真核生物外的其余真核生物的主要帽子形式。在有些真核生物中，在第二个核苷酸的 2′-O 位上还可以再产生甲基化，构成帽子 2(包括帽子 1 部分)，其符号为 $m^7GpppXmpYm$，帽子 2 (2 个甲基化)一般只占有帽子的mRNA的 10%～15%以下。3 种帽子的结构见图 7-18，所有帽子结构皆含 7 甲鸟苷酸(框示)，通过焦磷酸链连于 5′ 端。帽子的功能还不完全清楚，但已知对mRNA 的识别、结合和稳定有利。

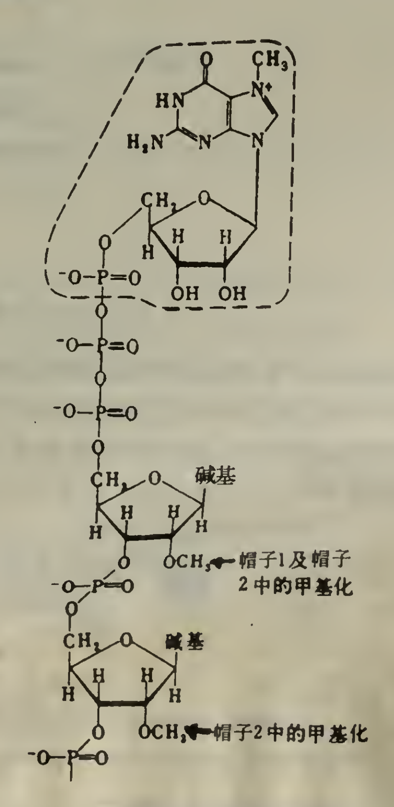


图 7-18 真核 mRNA 5′端的帽子结构
帽子 O 核糖无甲基化。

多聚(A)尾巴。有多聚(A)尾巴的 mRNA 写作 poly(A)$^+$，反之为 poly(A)$^-$。poly(A)是在转录后由 RNA 末端腺苷酸转移酶催化一个一个地加上去的。RNA 末端腺苷酸转移酶(RNA terminal riboadenylate transferase)的催化反应如下：

$$多聚核糖核苷酸 + nATP \xrightarrow[或 Mn^{2+}]{Mg^{2+}} 多聚核糖核苷酸(A)_n + nPPi$$

poly(A)的功能尚不清楚，但知 poly(A)与蛋白质结合，构成 mRNP，可能增加 mRNA 的稳定性。

(二)rRNA 转录后的加工

真核生物有 4 种 rRNA，即 5.8 S rRNA，18 S rRNA，28 S rRNA 和 5 S rRNA。前三者的基因组成一个转录元，产生 45 S 的前体 RNA。由于真核生物 rRNA 的加工过程比较缓慢，其中间产物可以分离出来，因此，真核生物 rRNA 的加工过程比较清楚。哺乳动物 45 S RNA 的加工有几种不同方式。例如，人类 45 S RNA 与小鼠的 45 S RNA 加工方式就不相同，其具体方式见图 7-19。

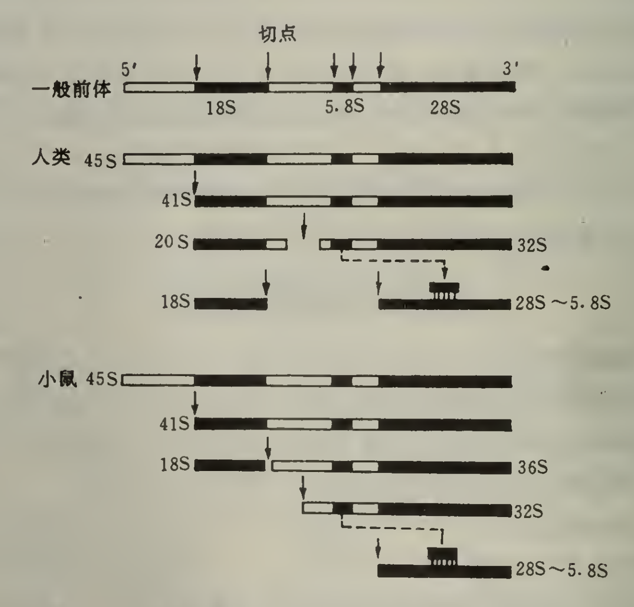


图 7-19　哺乳动物 45 S rRNA 前体的转录后处理

近年来发现，哺乳动物的原始转录本并不是 45 S，而是 47 S。由于 3′端部分很快进行转录处理，5′端也除掉一小部分故变为 45 S。

关于 5 S rRNA，与 tRNA 一起由 RNA 聚合酶Ⅲ转录。5 S rRNA 的基因含 120 bp，处在另一个转录单位中，这个单位含有 120 bp 的转录区和 600 bp 的非转录片段(见图 7-20)。



图 7-20　爪蟾 5 S rRNA 的基因结构

只有黑色区域被转录，长达 120 个单位。每个黑色区域和虚线之间有一个片段，其碱基顺序与 5S rRNA 很相近，可奇怪的是它并不被转录。

(三)tRNA 的转录后加工

其转录后的加工内容大致为：第一，切除 tRNA 前体两端多余序列。tRNA 前体 5′端较成

熟 tRNA 通常多几个到 10 个核苷酸，去除这些核苷酸是由高度专一的 RNase P（即 tRNA 5′端成熟酶）催化的。tRNA 3′端的切除由两个酶催化，即 tRNA3′端成熟核酸内切酶和核酸外切酶。第二是末端的添加，即在 3′添加 CCA 序列，此一步骤由 tRNA 核苷酰转移酶（tRNA nucleotidyl transferase）催化。第三，修饰。tRNA 修饰碱基很多，主要为甲基化修饰占被修饰碱基的一半以上。其它如 ψ 碱基（U 异化产生），Y 碱基（tRNA 特定位置上的 G37 转变而来），I 和 Q 主要系置换产生（前者是次黄嘌呤置换了前体 tRNA 上 34 位的腺嘌呤，后者是 Q 碱基置换了前体 tRNA34 位上的鸟嘌呤获得）。tRNA 转录后的加工参见图 7-21。

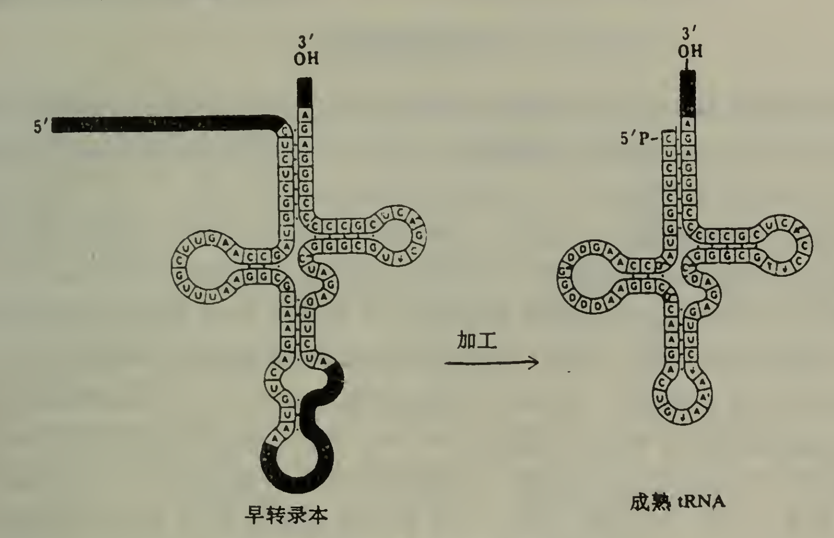


图 7-21　酵母酪氨酸 tRNA 前体的加工

一个 14 核苷酸的间隔顺序（标本）被去除，若干个碱基被修饰。通过断裂 5′-末端的领先段及在 3′-末端加上 CCA 顺序，由一个含 108 核苷酸的初级转录本形成了含 92 核苷酸的产物。

（四）居间序列的切除

居间序列（intervening sequence，IVS）或内含子（intron），在 RNA 的转录后加工中要除去，切除后要进行剪接，这是所有的 RNA 在转录后加工中必须进行的一个重要步骤。

RNA 剪接机制的研究，是 80 年代的热点之一，它不仅解决不连续基因转录产物的剪接问题，而且对于了解不连续基因的起源甚至整个生命的起源与进化都是有力的推动。核酸分子的催化功能的发现使人们有了对核酶的认识，近几年人们对于 RNA 剪接机制研究的进展主要内容如下。

不连续基因的居间序列又称内含子，被内含子隔开的基因序列称外显子（exon）。一个基因的外显子和内含子都转录在一条原初转录本 RNA 分子中，然后把内含子切除，把外显子连接起来，才能产生成熟的 RNA 分子。这个过程就叫 RNA 的剪接。内含子上游方向的一个拼接点称为 5′拼接点或左拼接点，内含子下游方向的一个拼接点称为 3′拼接点或右拼接点。

通过研究说明，内含子有各种类型。第一类内含子有二级结构，构成此二级结构的是 4 个 10～12 碱基的保守序列，线粒体与核基因属于此种类型。第二类内含子是不具备上述二级结构和保守序列的，如酵母线粒体细胞色素 c 氧化酶亚基 *a* 和亚基 *b* 基因的内含子 aI 1，aI 2，aI 5 和 bI 1 等。另一类内含子（也可能称为第三类）是核基因 mRNA 前体中的内含子。还有一类是 tRNA 基因的内含子。tRNA 前体的剪接不同于前 3 类内含子的剪接，前 3 类内含子的剪

接有共同特点。关于剪接的情况将在第十三章中详述。

在剪接RNA时，有两个磷酸二酯键破坏，同时形成一个新的磷酸二酯键。核酸内切酶与连接酶活性可能处于同一复合体上，剪接时协调进行。各类RNA内含子的切除及拼接皆有不同的特点，将在第十二章讨论。

第九节　RNA的自我剪切与催化

一、核酶的发现

真核生物中的核糖RNA分子与原核生物相似，也是靠初级转录本的裂解而形成的。Cech用四膜虫(*Tetrahymena*，一种原生动物)做的一个RNA剪切试验得到了意想不到的结果。从四膜虫的一个6.4 kb前体去掉一个414核苷酸内含子，从而得到一个26 S rRNA分子。原来他们是想将细胞提取液加到RNA前体中以测定剪切所需的蛋白质。很奇怪，他们发现在只含有RNA前体及三磷酸核苷的对照中并无蛋白质，却也发生了RNA的剪切(splicing)，为了避免RNA前体含有某种蛋白质，他们采用重组DNA法制备RNA前体，用纯化的编码这种前体的DNA在体外转录成RNA前体，以便进行剪切反应，结果也和上面所讲的一样，在核苷酸存在下，RNA自己将自己剪切。这个惊人的试验证明：一种RNA分子具有高度专一的催化活性，并且在无蛋白质情况下能进行自我剪切(self-splicing)。原来在此反应中要含有核苷酸，因为据认为需要ATP或GTP作为能源。后来证明所需要的辅助因子为一鸟苷(guanosine)单元，可以是鸟苷，GMP，GDP或GTP。G并非作为能源而是作为一攻击基团(attacking group)。G结合到RNA中，然后攻击5′剪切部位，与内含子的5′-末端形成一磷酸二酯键。这种转酯化作用(transesterification)反应在上游外显子末端生成一3′-OH。然后，3′剪切部位被此新生成的3′-OH基团攻击。这第二次转酯化反应将两个外显子连接起来并导致414nt内含子的释放。这个试验可以用图(图7-22)说明。在此反应中进行两种自我剪切，内含子的3′-OH攻击5′端的磷酸二酯键，形成一环状结构。在第一步中含有G的一个15-核苷酸片段掺入进去。所形成的399核苷酸环打开，形成一个线型分子，然后失去4-核苷酸片段，形成一个环状

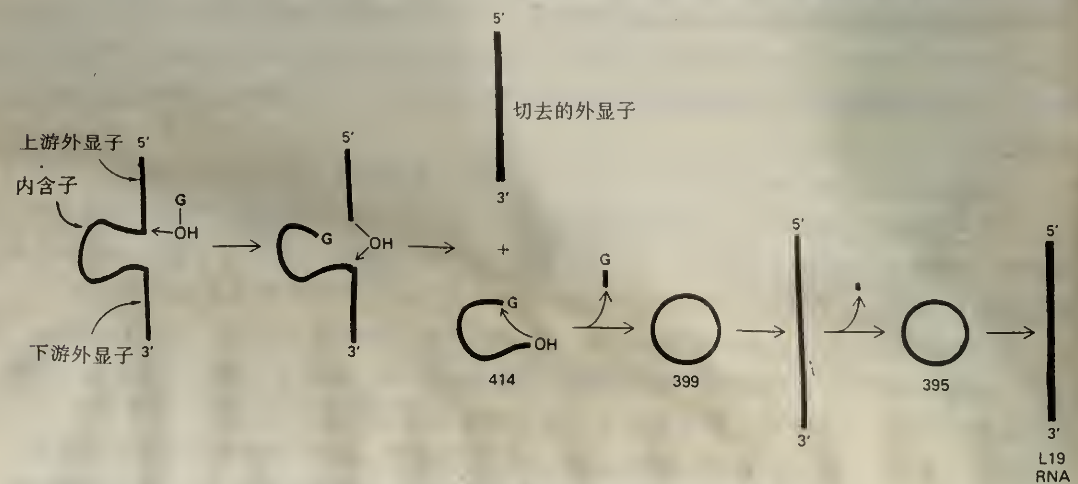


图7-22　四膜虫的核糖体RNA前体的自我剪切

结构，最后此环状物打开成一线型 RNA，称为 L 19 RNA（因为此 RNA 失去 19 个居间序列）。

RNA 自我剪切要依靠 rRNA 前体结构的完整。内含子的大部分对自我剪切都是必要的，rRNA 含有由许多双股螺旋茎部及套圈（loop）形成的折叠结构。折叠的 RNA 含有较弱的 GU 碱基对和较强的 AU 和 GC 碱基对，它们三者的亲和力之比为 1 ∶ 100 ∶ 1 000。RNA 前体含有一个专一的与 G 结合的口袋，在两个外显子之间形成一个磷酸二酯键，而内含子则以线型分子被释放出去。在此 rRNA 前体中键的形成与断裂的自我催化也是具有高度立体化学专一性的，与蛋白质酶类的催化作用相似。

二、L19 RNA 具有核酸酶及聚合酶性质

Cech 曾设想 L19 RNA 可能作用于外部底物上，实验证明确实如此。L19 RNA 能催化五胞嘧啶苷酸（pentacytidylate，C_5）成为更长或更短的寡聚体，见图 7-23。

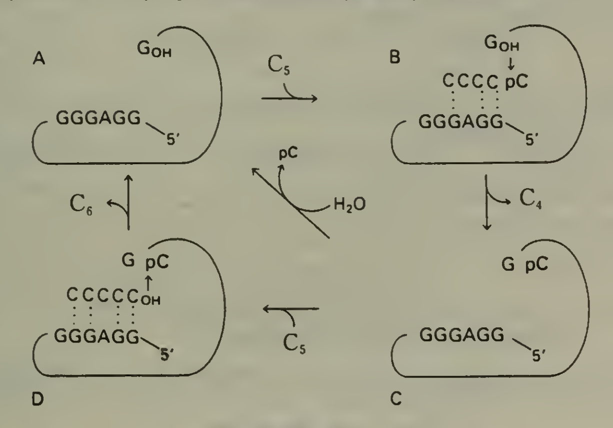


图 7-23　L19 RNA 的催化机理

A. 酶本身；B. C_5 与酶以氢键连接；pC 末端以共价键连于酶的 G 上；
C. GpC 水解，使酶复原；D. 第二个 C_5 分子攻击 pC 生成 C_6 分子。

由图 7-23 可见 L19 RNA 确实是一个真酶，它既是一个核酸酶（nuclease）也是一个聚合酶（polymerase）。这个酶（ribozyme，暂译为核酶）水解 C_5 的速率是未催化反应速率的 10^{10} 倍，这表明 RNA 分子具有很大的催化力。这个发现意味着 RNA 在生物进化的初期可以在没有蛋白质参与的情况下进行自我复制。

核酶究竟是如何将其底物切割呢？根据用一 13-聚体 RNA 做底物的研究结果提出下列切割方式：

G_{OH}

5′-GGCCCUCU(P)AAAAA-3′　13- 聚体底物

3′-GGGAGG-5′

三、只含 A3 核苷酸的锤头 RNA 具有催化活性

具有催化活性的 RNA 分子至少有多大呢？植物病源的 RNA 的研究表明核酶比 L19 RNA 要小得多。侵染植物的类病毒（viroid）为单股的环状 RNA 分子，约 300 nt 长，菲律宾椰子的一种病毒就是这种小 RNA 病毒。植物不含任何酶能裂解类病毒 RNA，类病毒也不编码

任何蛋白质，类病毒本身进行自我裂解(self-cleavage)。

比较核酶切割部位邻近的核苷酸序列，我们可以看到一个锤头(hammerhead)二级结构，它由一个不配对的中心辐射出3个螺旋组成(见图7-24)。根据这个模型，用化学方法可以合成一个19 nt RNA及一个24 nt RNA，形成锤头RNA。合成的锤头19 nt RNA是酶(核酶)而24 nt RNA则为底物。已经用许多种24 nt RNA加到19 nt RNA中证明核酶的作用。现在已经有可能设计合成核酶，使之在特定的部位切割靶RNA(target RNA)，以便阻止病原物的RNA病毒的基因或基因组的表达，达到治疗RNA病毒病的目的。Cech的出色研究成果获得了1989年诺贝尔化学奖。

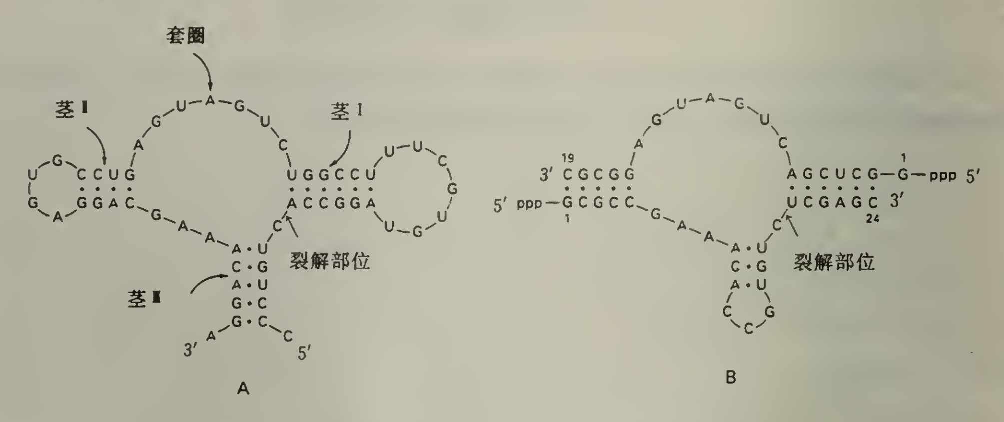


图7-24 锤头RNA的结构及其切割

A：烟草环斑病毒卫星RNA的自我切割序列；

B：人工合成的锤头催化RNA，1～19序列为核酶，1～24序列为底物。

参考文献

1. Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. and Weiner, A. M. Molecular Biology of the Gene. Vol I (4th ed.). Benjamin/Cummings, Menlo Park: 1987.
2. Lewin, I. V., Genes V. Oxford University Press, London: 1995.
3. Stryer, L. Biochemistry (4th ed.). W. H. Freeman and Company, New York: 1995.
4. Freifelder, D. Molecular Biology (2nd ed.). Johns and Bartlett Publisher, Boston: 1987.

(陈祖洁)

第八章　遗 传 密 码

在 1954 年，Gamow 首先提出遗传密码(genetic code)的问题。遗传密码是指 DNA(或其转录本 mRNA)中碱基序列和蛋白质中氨基酸序列之间的相互关系。这便存在由几个核苷酸决定一个氨基酸的问题。由于蛋白质含有 20 种氨基酸，而核苷酸只有 4 种，所以，至少要由 3 个核苷酸组成一个密码，这样便可以有 $4\times4\times4=64$ 种密码，以满足要求。Brenner 和 Crick 在 1961 年的实验证明，加入或减少 1 或 2 个核苷酸便导致形成不正常的蛋白质，但加入或减少 3 个核苷酸则生成的蛋白质常常具有完全的活性。因此，他们认为遗传密码是由 3 个核苷酸组成的，这也为以后的试验所证明。

但至 1960 年，虽然 RNA 如何参与蛋白质合成的总轮廓已经清楚，却仍未了解遗传密码的细节。破译遗传密码的最直接而又令人信服的途径是测定一个基因的核苷酸顺序及与其对应的多肽的氨基酸顺序，然后进行比较。但那时只有测定氨基酸顺序的有效方法，而没有测定核苷酸顺序的成熟方法。1961 年 Nirenberg 等在大肠杆菌提取液中，借助于添加合成的多聚尿嘧啶核苷酸而合成出只含 1 种氨基酸(苯丙氨酸)的多肽，这说明 polyU 中含有苯丙氨酸的密码，从而部分地破译了遗传密码。随后经过科学家们的多年努力，终于对遗传密码有了较全面地了解。

本章讨论遗传密码的破译及其特性，基因突变及基抑制或校正，以及重叠基因等内容。

第一节　遗传密码的破译

一、蛋白质的体外合成

对密码子的破译起很大作用的是蛋白质体外合成的试验。用活跃地进行蛋白质合成的大肠杆菌制备无细胞提取液，并用放射性同位素标记(^{3}H，^{14}C，^{35}S)的氨基酸以研究氨基酸掺入到蛋白质中的情况(图 8-1)。(a)在冷冻(0℃)中离心收集迅速生长的大肠杆菌细胞，破碎，获得细胞液，加入脱氧核糖核酸酶以破坏 DNA；(b)细胞液离心(0℃)除去细胞壁和细胞膜碎片。收集沉降较慢的组分(多核糖体、游离核糖体、游离 mRNA 和 tRNA、酶)于试管中；每管加入 ATP，GTP 和放射性氨基酸，将试管放在 37℃下保温不同时间；(c)加入酸沉淀蛋白质；游离氨基酸仍在溶液中。洗涤沉淀，收集，放入计数器计数，从沉淀的放射活性测量出氨基酸掺入(蛋白质合成)量。用这种方法证明蛋白质的合成需要 3 种 RNA，即 tRNA，rRNA 和 mRNA 的参与。但是，这个体外合成只进行几分钟便逐渐减慢以至停止。这是由于在提取液中存在降解 mRNA 的酶，使 mRNA 分解，如果加入新的 mRNA 于已停止合成蛋白质的提取液，又可以重新开始合成。这不但证明蛋白质的合成需要 mRNA 作为模板，而且用已停止合成蛋白质的提取液，加入不同的 mRNA，便可以研究出 mRNA 是如何编码蛋白质的合成了。

但是，在体外合成的蛋白质，是否和在体内合成的天然蛋白质一样呢？在 1962 年，用大肠

杆菌提取液进行体外合成试验，加入细菌病毒 f2 的 RNA，结果合成了一个与天然的 f2 外壳蛋白完全相同的蛋白质，其氨基酸顺序完全一样。这便证明在体外条件下，可以准确地按照 mRNA 的遗传信息合成相应的蛋白质。

在随后的 10 年中，已成功地用无细胞系统合成多种其它的病毒和细菌蛋白。例如，T4 溶菌酶，它是一个分子量为20 kD的分解细菌细胞壁的酶。用无细胞系统合成的 T4 溶菌酶具有酶促活性，这便进一步证明在体外正确翻译的可能性。以后又合成了有酶活性的 β-半乳糖苷酶（一种含 1 300 个氨基酸的蛋白质）。随后，又在试管内合成了多种重要的哺乳动物蛋白（如血红蛋白、轻链和重链免疫球蛋白、肌动蛋白、胶原蛋白、肌球蛋白），证明高等生物细胞蛋白也能在体外合成。

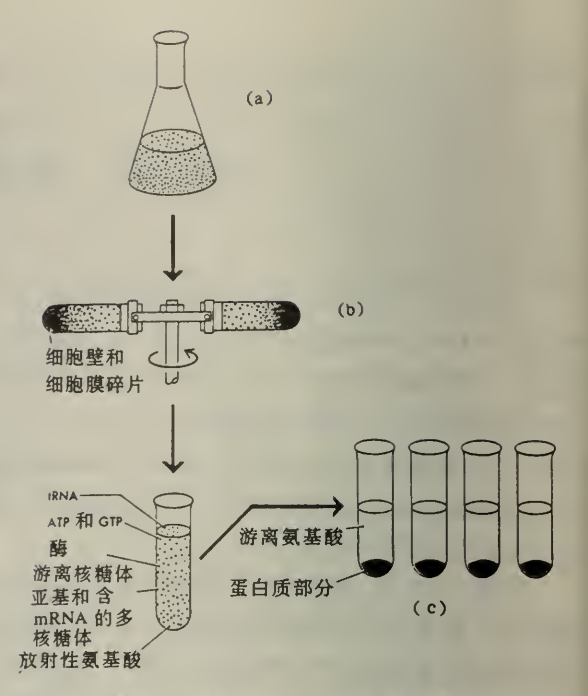


图 8-1　体外蛋白质合成试验步骤

二、用人工合成的 mRNA 进行蛋白质合成

用人工合成的已知其碱基顺序的 mRNA 进行体外蛋白质合成，观察氨基酸掺入情况，便可以了解编码各种不同氨基酸的密码子。

在 1955 年 Grunberg-Manago 和 Ochoa 从细菌中分离出多核苷酸磷酸化酶（polynucleotide phosphorylase），催化下列反应：

$$\text{RNA} + \text{Ⓟ} \rightleftharpoons \text{核糖核苷}-\text{Ⓟ}\sim\text{Ⓟ}$$

在核苷二磷酸开始浓度很高时，此酶可以催化核苷酸间形成 3′-5′ 磷酸二酯键，生成 RNA 分子：

$$\underset{\text{多聚 }A_n}{5'\cdots \text{ApApA}_{\text{OH}}3'} + \text{ppA} \rightleftharpoons \underset{\text{多聚 }A_{n+1}}{\cdots \text{ApApApA}_{\text{OH}}} + \text{Ⓟ}$$

用多核苷酸磷酸化酶合成 RNA 并不需要模板 DNA 或 RNA；人工合成产物的碱基成分完全随加入反应混合物中的不同核糖核苷二磷酸的比例而异。例如，当只加入腺苷二磷酸时，生成的 RNA 只含腺苷酸，称为多腺苷酸或多聚 A。同样也可合成多聚 U、多聚 C、多聚 G。加入两个或多个不同的核苷二磷酸，则生成混合共聚物，如多聚 AU、多聚 AC、多聚 CU、多聚 AGCU 等。这些混合共聚物的碱基顺序近似于随机排列，而且决定于反应物的相对浓度。例如，用 A∶U＝2∶1 混合物，则生成的多 AU 的碱基顺序可能为 UAAUAUAAAUAAUAAAAUAUU……。

在 1961 年，Nirenberg 和 Matthaei 用上述合成了多聚尿苷酸（多聚 U）。他们发现，在由大肠杆菌制备的无细胞蛋白质合成系统中加入多聚 U 和各种氨基酸，结果生成只含有苯丙氨酸的多肽链。根据三联体密码的模型，便可知苯丙氨酸的密码子应是 UUU。随后其他研究者亦

用同样技术证明赖氨酸的密码子是 AAA，脯氨酸的密码子是 CCC。但用这方法未能破译 GGG 所编码的氨基酸，因为多聚 G 中的鸟苷酸残基以氢键牢固地彼此连结成三股螺旋(triple helix)，因而防碍其与核糖体结合。

但是用这方法却不能破译含两种以上的碱基的密码子。例如，多聚 AC 分子可以含有 8 种不同的密码子：CCC，CCA，CAC，ACC，CAA，ACA，AAC，AAA。除已知 CCC 和 AAA 分别编码脯氨酸和赖氨酸外，其它密码子则无从破译。

当用多聚 AC 进行蛋白质合成时，除了脯氨酸和赖氨酸外，还能使天冬酰胺、谷氨酰胺、组氨酸、苏氨酸掺入到蛋白质中去。这些氨基酸的掺入比例，随 A/C 比率而异。如果在多聚 AC 中 A 的量大大超过 C，则天冬酰胺的掺入大大多于组氨酸。这样便可推断，天冬酰胺是由 2A 和 1C 编码的，而组氨酸是由 2C 和 1A 编码的。用其它共聚物进行类似试验，亦可推断出其它的密码子的碱基组成。但这些试验仍未能确定一个密码子内的碱基顺序，如组氨酸的密码子可能是 CCA，也可能是 CAC 或 ACC。

三、用 tRNA 结合法破译密码子

在 1964 年，Nirenberg 发现，即使没有蛋白质合成所需的全部因子存在，特异的氨酰-tRNA分子也可以与核糖体-mRNA 复合物结合。例如，多聚 U 与核糖体形成的复合物只与苯丙氨酰-tRNA 结合。而多聚 C 与核糖体的复合物则只与脯氨酰-tRNA 结合。而且，这个特异结合并不需要长的 mRNA 分子，只要一个短的三核苷酸就足够了。例如，加入三核苷酸 UUU 可与苯丙氨酰-tRNA 结合，而 AAA 则可与赖氨酰-tRNA 特异地结合于核糖体。用已知碱基顺序的三核苷酸进行试验，便可以测知各种不同氨基酸的密码子了。表 8-1 列出一些三核苷酸-核糖体复合物所结合的氨酰-tRNA。但是，用这个方法还未能确定全部氨基酸的密码子，因为有些三核苷酸的结合率十分低，以致无法确定哪一个密码子对应于哪个氨基酸。而且，用这方法也把一些密码定错了，如表 8-1 中的 UCU 实际上是丝氨酸而不是缬氨酸的密码子。

表 8-1　氨酰 tRNA 结合于三核苷酸-核糖体复合物

三核苷酸	结合的 AA-tRNA
5′-UUU-3′，UUC	苯丙氨酸
UUA，UUG，CUU，CUC，CUA，CUG	亮氨酸
AUU，AUC，AUA	异亮氨酸
AUG	甲硫氨酸
GUU，GUC，GUA，GUG，UCU*	缬氨酸
UCU，UCC，UCA，UCG	丝氨酸
CCU，CCC，CCA，CCG	脯氨酸
AAA，AAG	赖氨酸
UGU，UGC	半胱氨酸
GAA，GAG	谷氨酸

*注意用这方法把这密码子定错了。

四、用重复共聚物确定密码子

在应用三核苷酸技术的同时，Khorana 用有机化学与酶学技术相结合的方法合成了已知顺序的含 2，3，4 种碱基的共聚物。这方法如图 8-2 所示。首先，用有机化学方法合成了含 9 个

碱基的两条互补的脱氧寡核苷酸链 d(TAC)$_3$ 和 d(GTA)$_3$，以此为模板，用 DNA 聚合酶 I 合成长链的 DNA。又以此为模板，在 RNA 聚合酶作用下，如果加入 GTP，UTP 和 ATP，便合成多聚(GUA)的 RNA 链(图 8-2 左)；如果加入的是 CTP，UTP，ATP，便合成多聚(UAC)链(图 8-2 右)。这样，便可以合成各种不同的重复共聚物，并用以研究密码子，例如，用重复顺序 CUCUCUCU……做模板，结果生成亮氨酸与丝氨酸交替排列的多肽，由此推知亮氨酸的密码子是 CUC，而丝氨酸的密码子是 UCU。同样，用重复顺序 UGUGUGUG……做模板，便得出半胱氨酸与缬氨酸交替排列的多肽，并推知半胱氨酸的密码子是 UGU，缬氨酸的密码子是 GUG。

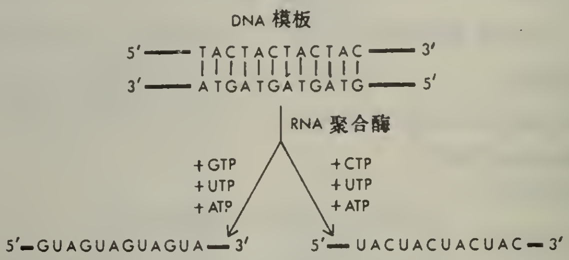


图 8-2　用重复顺序双链 DNA 为模板合成具有不同重复顺序的 RNA 链

用同样的方法，也可以测知含 3 种碱基重复顺序所编码的多肽。例如，用多聚 AAG(AAGAAGAAG……)得出 3 种多肽：多聚赖氨酸(密码子是 AAG)、多聚精氨酸(密码子是 AGA)和多聚谷氨酸(密码子是 GAA)。同样，多聚 AUC 则是多聚异亮氨酸(密码子 AUC)、多聚丝氨酸(密码子 UCA)、多聚组氨酸(密码子 CAU)的模板(表 8-2)。甚至用 4 个碱基重复顺序，亦可以测知不同氨基酸的密码子。例如，用多聚(UAUC)，则合成的多肽为酪氨酸-亮氨酸-丝氨酸-异亮氨酸重复顺序的多肽：

UAU ⋮ CUA ⋮ UCU ⋮ AUC ⋮ UAU ⋮ CUA ⋮ UCU ⋮ AUC
酪　亮　丝　异亮　酪　亮　丝　异亮

表 8-2　用 2 个或 3 个核苷酸构建的重复共聚物确定的密码子

共聚物	识别的密码子	氨基酸掺入或生成的多肽	确定的密码子
(CU)$_n$	CUC ⋮ UCU ⋮ CUC…	亮氨酸	5′-CUC-3′
		丝氨酸	UCU
(UG)$_n$	UGU ⋮ GUG ⋮ UGU…	半胱氨酸	UGU
		缬氨酸	GUG
(AC)$_n$	ACA ⋮ CAC ⋮ ACA…	苏氨酸	ACA
		组氨酸	CAC
(AG)$_n$	AGA ⋮ GAG ⋮ AGA…	精氨酸	AGA
		谷氨酰胺	GAG
(AUC)$_n$	AUC ⋮ AUC ⋮ AUC…	多聚异亮氨酸	5′-AUC-3′
	UCA ⋮ UCA ⋮ UCA…	多聚丝氨酸	UCA
	CAU ⋮ CAU ⋮ CAU…	多聚组氨酸	CAU

这样，用 Nirenberg 的三核苷酸结合技术，和 Khorana 的重复顺序技术，终于在 1966 年，将遗传密码完全破译了。表 8-3 列出遗传密码，其中 61 个是编码各种氨基酸的，另 3 个是终止密码子(详见下文)。

表 8-3 遗传密码

第一位（5′端）	第二位 U	第二位 C	第二位 A	第二位 G	第三位（3′端）
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U
U	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C
U	UUA Leu	UCA Ser	UAA 终止	UGA 终止	A
U	UUG Leu	UCG Ser	UAG 终止	UGG Trp	G
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U
C	CUC Leu	CCC Pro	CAC His	CGC Arg	C
C	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A
C	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U
A	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C
A	AUA Met	ACA Thr	AAA Lys	AGA Arg	A
A	AUG* Met	ACG Thr	AAG Lys	AGG Arg	G
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U
G	GUC Val	GCC Ala	GAC Asp	GGC Gly	C
G	GUA Val	GCA Ala	GAA Glu	GGA Gly	A
G	GUG* Val	GCG Ala	GAG Glu	GGG Gly	G

*亦作为甲酰甲硫氨酰-$tRNA^{Met}$ 的起始密码子。GUG 同时编码缬氨酸和甲硫氨酸，是双关密码子(ambiguous codon)。

由于 Nirenberg 和 Khorana 2 人在破译遗传密码研究工作中的重大贡献，他们 2 人共同获得 1968 年的诺贝尔奖。

审察表 8-3 的遗传密码，可以看出下列特点：

(1)除色氨酸和甲硫氨酸只有 1 个密码子外，其它 18 种氨基酸均有 1 个以上的密码子，如苯丙氨酸、酪氨酸、组氨酸、谷氨酰胺、谷氨酸、天冬酰胺、天冬氨酸、赖氨酸、半胱氨酸各有 2 个密码子；异亮氨酸有 3 个密码子；缬氨酸、脯氨酸、苏氨酸、丙氨酸、甘氨酸各有 4 个密码子；亮氨酸、精氨酸、丝氨酸各有 6 个密码子。

(2)许多氨基酸的密码子均在同一方框内，即第一和第二个碱基相同，只有第三个碱基不同，这称为"非混合型密码子族(unmixed codon family)"，如上文列出的有 4 个密码子的氨基酸。有些方框内有两种氨基酸的密码子，这些称为混合型密码子族(mixed codon family)。它们的第一、二个碱基均相同，第三个嘌呤或嘧啶分别编码不同氨基酸。有 6 个密码子的氨基酸的密码子分布在不同方框内，它们的第一或第一和第二个碱基不同，如亮氨酸的密码子第一个碱基为 U 或 C，精氨酸的第一个碱基为 C 或 A，丝氨酸密码子的第一个碱基为 U 或 A，第二个碱基为 C 或 G。

(3)甲硫氨酸的密码子 AUG 同时也作为起始密码子(见下文)。

(4)性质相近的氨基酸的密码子分布在相近位置，如表中第一纵行(第二碱基为 U)的苯丙氨酸、亮氨酸、异亮氨酸、甲硫氨酸、缬氨酸，以及第二纵行(第二碱基为 C)的脯氨酸和丙氨酸均为疏水氨基酸；第三和第四纵行的组氨酸、赖氨酸、天冬氨酸、谷氨酸和精氨酸均为带电荷的氨基酸；第二、三、四纵行的丝氨酸、苏氨酸、天冬酰胺、谷氨酰胺、半胱氨酸、甘氨酸则为不带电荷的氨基酸。遗传密码的这种特性，可使发生突变时，对蛋白质的性质不会产生太大影响(见下文)。

(5)在 mRNA 模板上的密码子是连续的，在前一个密码子与后一个密码子之间没有间隔，即没有一个间断的讯号。因此，在进行翻译时，解读的框架(frame)决定于起始的碱基。如果在

解读开始时移动了1个或2个碱基，或者核糖体在mRNA链上移动时偶然跳越过1个核苷酸，便会生成不完全的蛋白质（见本章第五节及第九章）。

第二节　密码的简并性与变偶假说

许多氨基酸都是由多个密码子编码的，这称为简并性(degeneracy)。例如，UUU和UUC编码苯丙氨酸，UCU，UCC，UCA，UCG，AGU和AGC编码丝氨酸。事实上，在头两个核苷酸相同时，第三个核苷酸则可以是胞嘧啶或尿嘧啶，也可以是腺嘌呤或鸟嘌呤，而仍是编码同一个氨基酸。但并不是所有的简并性都是头两个核苷酸相同。例如，亮氨酸可由UUA和UUG编码，也可由CUU，CUC，CUA，CUG编码，它们分别由两种tRNA与之结合（图8-3）。密码子的简并性，特别是第三位的胞嘧啶和尿嘧啶或鸟嘌呤和腺嘌呤的简并性常常等同，这说明为什么在不同生物的DNA中的AT/GC比率会有很大的变异，而其蛋白质的氨基酸相对比例却没有很大的变化。

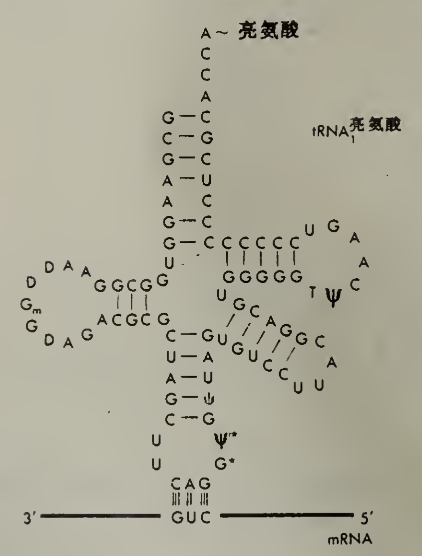


图8-3　大肠杆菌的两种不同的携带亮氨酸的tRNA分子

* 表示性质未明的修饰。

由于携带氨基酸的tRNA是以其反密码子与mRNA上的密码子碱基配对的，那么看来必须有61种tRNA与61个密码子配对。但人们发现有些tRNA（如丙氨酰-$tRNA^{Ala}$）可以和几种密码子配对。而且还发现有些tRNA的反密码子含有稀有碱基次黄嘌呤（Ⅰ），它是由腺嘌呤的6位C上脱氨生成6-酮基而形成的：

次黄嘌呤可以与A，U或C配对。在1966年，Crick提出“变偶假说”(wobble hypothesis)，用以解释上述现象。这个假说包括下述内容：

(1)mRNA上的密码子的第一、第二个碱基与tRNA上的反密码子相应的碱基形成强的配对；密码的专一性，主要是由于这两个碱基对的作用。

(2)反密码子的第一个碱基（以5′→3′方向，与密码子的第三个碱基配对）决定一个tRNA所能解读的密码子数目。当反密码子的第一个碱基是C或A时，则只能和一个密码子结合。但当反密码子上的第一个碱基是U或G时，则可以和两个密码子结

合，即 U 可以和 A 或 G 配对，G 可以和 C 或 U 配对。而当反密码子的第一个碱基是 I 时，便可以和 3 个密码子结合，即 I 可以和 A，U 或 C 配对。如表 8-4 所示：

(3)当一种氨基酸是由几个不同的密码子编码时，如果密码子的头 2 个碱基的任一个是不同的，便必须有不同的 tRNA。

(4)这样，便要求至少要有 32 种 tRNA，来与 61 个密码子相结合。

表 8-4 根据变偶假说的碱基配对

反密码子的第一个碱基	密码子的第三个碱基
G	U 或 C
C	G
A	U
U	A 或 G
I	A，U 或 C

密码子与反密码子相互作用的复杂情况，表明密码子的头 2 个碱基在决定其专一性中是起主要作用的。第三个碱基(变偶碱基)对专一性也起一定作用，但因其与反密码子中的相应碱基结合较弱，在蛋白质合成时这种 tRNA 便能以较快速度与密码子解离，这便有利于加快蛋白质的合成。

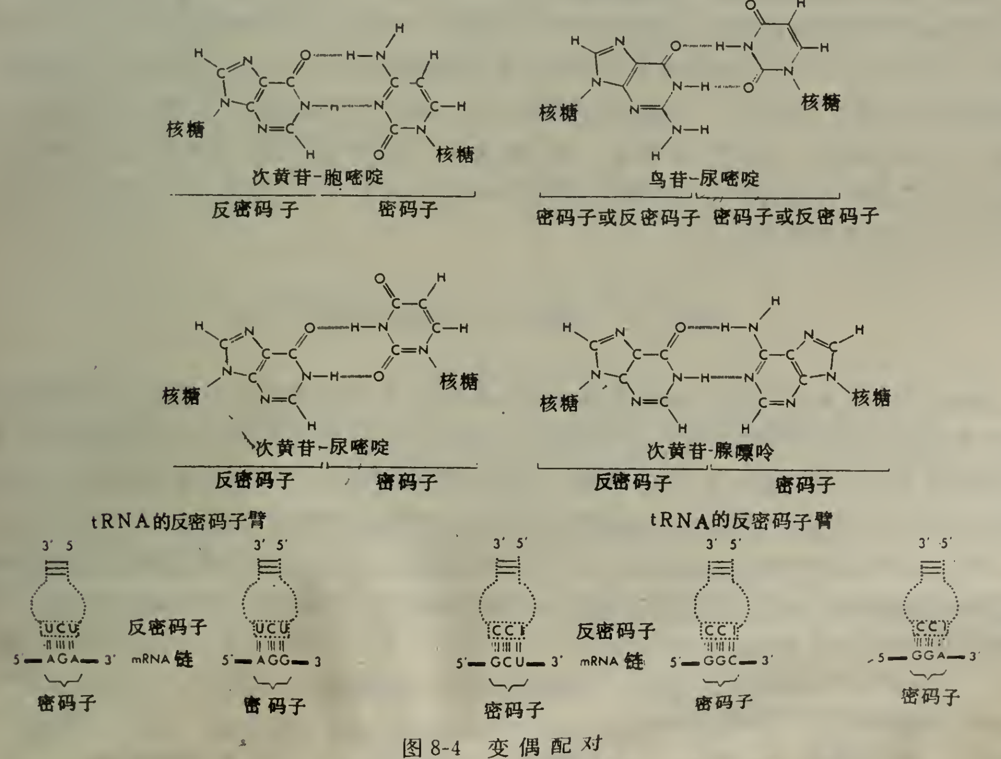


图 8-4 变偶配对

上图：变偶作用能使反密码子的碱基与非标准配对的碱基形成氢键。下图左：反密码子第一位的 U 可与 A 或 G 配对；下图右：反密码子第一位的 I 可与 U，C 或 A 配对。

这些可以发生配对的碱基的核糖-核糖之间的距离，与标准 AU 或 GC 配对的距离相近似。而嘌呤与嘌呤或嘧啶与嘧啶之间的配对，均使核糖-核糖之间的距离过长或过短，因而不能发生配对。图 8-4 列出根据变偶假说可发生配对的例子。从图 8-4 看出，变偶配对的核糖-核糖距离与标准的 AU 或 GC 配对相接近。

自1966年以后，差不多所有的试验结果均支持变偶假说。例如，根据变偶假说，有6个密码子的丝氨酸便至少有3种tRNA，事实也证明如此。精氨酸和亮氨酸也有6个密码子，它们的第一个碱基不完全相同，这些氨基酸也有不同的tRNA，用以与第一或第二个碱基不同的密码子结合。

根据近年确定的tRNA三维结构，反密码子的3个碱基以及其3′端以后的2个碱基均是指向差不多同一的方向的，其构象主要决定于碱基的平面之间的堆叠相互作用。这样，反密码子的5′端碱基便位于堆叠下方，因而在与密码子3′端的碱基配对时便不那么受限制而可以发生摇摆（“wobble”原意为“摇摆”，变偶假说也有人译作“摇摆假说”。）而反密码子的3′端碱基则位于堆叠的中央，而且邻近的碱基又是体积较大的修饰嘌呤碱，因而限制了它的摆动。

然而，也发现一些现象不符合变偶假说。例如，缬氨酸的tRNA中的任何一种均可以和缬氨酸的4个密码子结合。而按照变偶假说，缬氨酸的3种tRNA中的2种反密码子GAC和UAC最多只能和2种密码子结合，第三种tRNA的反密码子IAC也只能和3种密码子结合，这便不能说明上述现象。

在原核生物和在真核生物的细胞质内，反密码子的5′位U通常是经过修饰的，这可以影响其与密码子可能配对的范围。在线粒体内发现约有一半的tRNA反密码子的5′位U是未经修饰的。这些tRNA可用以解读非混合密码子族（即密码子XYU，XYC，XYG，XYA编码同一种氨基酸）的全部4个密码子。但在混合密码子族中，密码子XYA和XYG则由同一种在反密码子的5′位具有修饰U的tRNA解读。这样看来，U的修饰显然是将其解读的专一性限于嘌呤碱。在混合密码子族的另2种末端为嘧啶的密码子（XYU，XYC）则由在反密码子5′位为G的tRNA与之按变偶假说配对。

第三节 密码子的使用频率

密码子具有简并性，即同一种氨基酸具有多个密码子，那么是否各个密码子均以相同或近似的几率被使用，或是其中有些优先被使用呢？例如，在具有富含AT的DNA的生物中，是否大多数密码子的第三位都会有U或A呢？研究发现，在噬菌体ΦX174的单链DNA基因组中含A24%，C22%，G23%，T31%。其密码子中的第三位是优先使用U的。

即使在有些生物中的碱基组成是比较均匀的，其密码子的使用也不是随机的。在大肠杆菌中即发现，有些密码子被反复使用，而有些则几乎不被使用。例如，在大肠杆菌核糖体蛋白质基因的1 209个密码子中，编码苏氨酸的4个密码子中的ACU被使用36次，ACC 26次，而ACA只有3次，ACG则完全没有使用。动物密码子也有类似情况。据统计，在2 244个密码子中，编码丙氨酸的GCU密码子有28个，GCC 38个，GCA 14个，GCG 6个。其它氨基酸密码子的使用也有类似情况。上述例子说明，编码同一个氨基酸的各个密码子的使用不是平均的。

在第十章将谈到，一种氨基酸有1个以上的tRNA。有时这些tRNA的碱基顺序虽然不同，其反密码子却是相同的。这些携带同一氨基酸的不同tRNA的量常有很大差异（可达11倍）。例如，在大肠杆菌中有一种主要的和一种次要的酪氨酰-tRNA，它们的反密码子均为3′-AUG-5′。这次要tRNA的功能还不清楚。

据研究，大肠杆菌mRNA中的密码子出现频率与相应的tRNA含量有高度相关性。图8-5表示大肠杆菌几个基因的tRNA使用频率与其含量的相关性。图的横坐标为26种tRNA的

相对量，以 $tRNA_1^{Leu}$ 为 1.0。根据这些 tRNA 的已知反密码子的碱基顺序，算出 4 种基因中相应的密码子出现频率。从图 8-5 中看出，高度表达的 *recA* 基因和核糖体蛋白基因，tRNA 量与使用频率呈高度相关，而 β-内酰胺酶(β-lactamase)基因和噬菌体 *ΦX*174 基因则各点很分散。在高等生物内也是如此，被主要的 tRNA 同工受体解读的，也就是那些使用频率最高的密码子。例如，酵母的醇脱氢酶同工酶 I 和磷酸甘油醛脱氢酶，它们的氨基酸的 96%只由 25 种密码子编码，其相应的 tRNA 也是十分丰富的。因此，高度表达的基因中，密码子的选择受控于可供使用的 tRNA 的含量，而且密码子的选择是使其翻译效率达到最佳状态，而稀有 tRNA 的密码子则可导致翻译减慢。

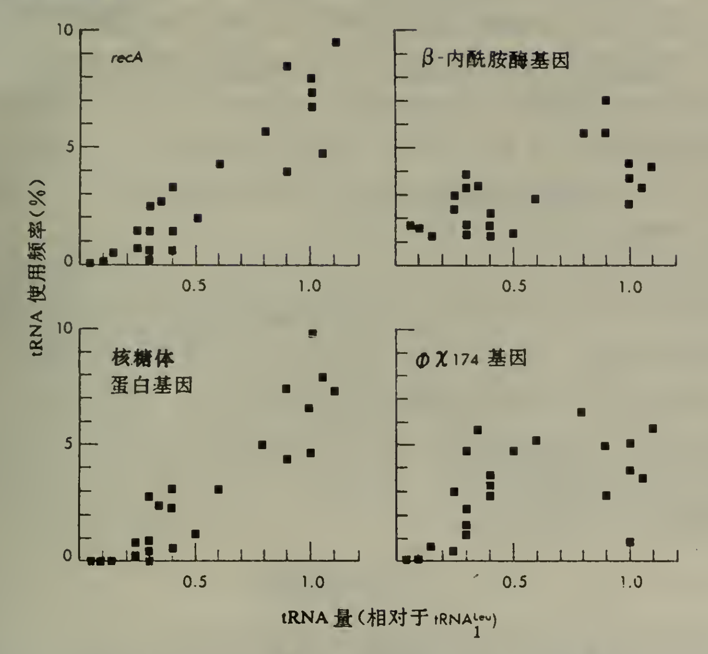


图 8-5　大肠杆菌中的 tRNA 量与 4 个基因中 tRNA 使用频率的相关关系
tRNA 以相对量表示，以 $tRNA_1^{Leu}$ 的量为 1.0。

这样，即使多个密码子可由变偶作用与一种 tRNA 配对，但通常是其中一个密码子被优先结合。密码子的优先结合有下列几种情况：

(1)反密码子的 5′位(变偶位)U 被修饰可使之优先与密码子第三位的 A 配对胜过与 G 配对。

(2)反密码子的 5′位的 I 与密码子的 U 和 C 配对优先于与 A 配对。

(3)当一个密码子的头 2 个碱基与反密码子形成 AU 配对时，则第三个碱基优先用 C(生成 GC 配对)而不用 U(生成 AU 对)，因前者较后者稳定。

根据上述几个规律，同时也根据细胞内各种 tRNA 的相对量，我们可以推测在一个细胞内的 mRNA 中的密码子使用情况。

第四节　起始密码子与终止密码子

一、起始密码子

密码子 AUG 是与 N-甲酰甲硫氨酰-tRNA 结合的密码子，在原核生物中启动蛋白质的合成，所以叫做起始密码子(initiation codon)。起先人们发现有一种专门携带 N-甲酰甲硫氨酸的 tRNA($tRNA_f^{Met}$)，因此推想，它会识别与甲硫氨酸的密码子 AUG 不同的另一种密码子。但是，将 $tRNA_f^{Met}$ 进行顺序分析，却发现它的反密码子与运载甲硫氨酸的 $tRNA_m^{Met}$ 一样，即 3′-UAC-5′。即是说，$tRNA_f^{Met}$ 和 $tRNA_m^{Met}$ 都是与 AUG 配对的。那么，如何区别这两种 tRNA 的配对呢？这是由蛋白质合成的因子决定的，因为只有 fMet-$tRNA_f^{Met}$ 才能和起始因子 IF2 结合以生成 30 S 起始复合物；也只有 Met-$tRNA_m^{Met}$ 才能与延长因子 EF-Tu 结合。这样，虽然它们共用一个密码子 AUG，但甲酰甲硫氨酸只能在多肽链合成起始时掺入，而甲硫氨酸则只能在多肽链延长时掺入到肽链内部。

随后又发现，fMet-$tRNA_f^{Met}$ 还可以与 GUG 密码子结合。在大肠杆菌内，GUG 作为起始密码子的频率只有 AUG 的 1/30。GUG 原是缬氨酸的密码子。GUG 与 $tRNA_f^{Met}$ 的反密码子 3′-UAC-5′配对，便意味着有一种新的变偶。因为在这里，不是密码子的第三个(3′端)碱基而是第一个(5′端)与反密码子发生变偶。对这种异常的变偶的一个可能的解释是，根据 $tRNA_f^{Met}$ 的顺序分析，邻近反密码子 3′端的碱基是未经修饰的腺嘌呤，而不是在差不多所有其它 tRNA 中都存在的体积巨大的烷基化衍生物。不仅如此，甚至 UUG 和 CUG 有时也具有起始密码子的作用，但其频率则比 GUG 更低。

二、终止密码子

密码子 UAA，UAG 和 UGA 并不编码任何氨基酸，因此，也称为无义密码子(nonsense codon)。但这个名称并不恰当，因为它们虽然不编码任何氨基酸，却起着终止肽链合成的作用，因此，称为终止密码子(termination codon)。UAA 也称为赭石型(ochre)，UAG 称为琥珀型(amber)，UGA 称为乳白型(opal)密码子。所有这 3 个密码子均是作为肽链终止的密码子，它们在蛋白质合成中起着终止肽链延长的作用。在第九章将谈到，有两个释放因子 RF1 和 RF2，它们分别识别 2 个终止密码子：RF1 识别 UAA 和 UAG，RF2 识别 UAA 和 UGA。RF1 和 RF2 均是蛋白质，这便表明，多核苷酸不仅可以和另一种多核苷酸相互作用，也可以和蛋白质起相互作用；也即是说，不仅碱基与碱基之间可以生成氢键而互相识别，也可以和蛋白质中的氨基酸生成氢键而被识别。

尚不清楚为什么需要 3 个终止密码子。据对大肠杆菌的一些基因所做的碱基顺序分析，多数基因是用 UAA 作为终止密码子。有些基因有 2 个甚至 3 个连续的终止密码子(图 8-6)。例如，噬菌体 *Qβ* 的外壳蛋白基因只有一个终止密码子，而噬菌体 R17 和 MS2 的外壳蛋白基因则有 2 个终止密码子。从图 8-6 可以看出，在有些 mRNA 中，2 个或 3 个终止密码子是连续的，但有的则分隔开。多个终止密码子的存在，对保证蛋白质合成的适时终止是很重要的。

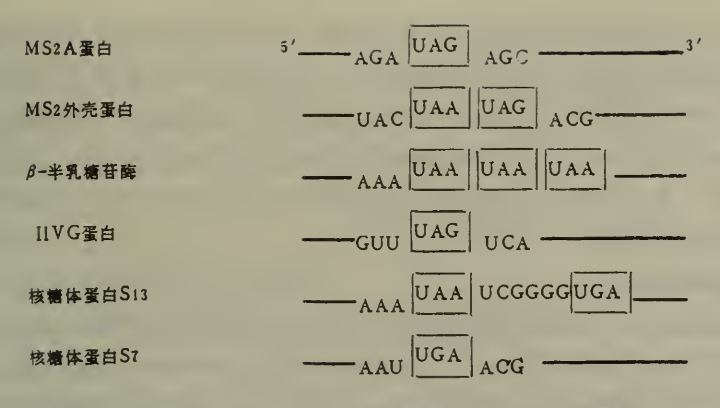


图 8-6 在大肠杆菌内几种 mRNA 编码区 3′端的核苷酸顺序
终止密码子以方框表示。

第五节 遗传密码的突变

当基因组的 DNA 的顺序发生变化时，便产生突变(mutation)。本节从遗传密码的角度，讨论突变发生的情况及其抑制或校正。

一、无义与错义突变

突变有两种情况：当一个密码子发生改变，使这个密码子编码出另一个"错误"的氨基酸，这便称为错义突变(missense mutation)。如果肽链终止无义密码子改变为有义密码子，或有义密码子改变为无义密码子，便称为无义突变(nonsense mutation)。

由于终止密码子只有 3 个，而编码氨基酸的密码子有 61 个，所以，如果只改变 1 个碱基(称为点突变，point mutation)，那么通常是发生错义突变，而无义突变出现的几率较少。由错义突变产生的蛋白质只改变 1 个氨基酸，仍具有原来的一些生物学活性。上文已指出，在遗传密码表上，性质相同的氨基酸的密码子是排在附近的，因此，发生错义突变对生成的蛋白质的性质往往无大影响。它们常能保持原来的蛋白质的某些活性。在高温下，错义蛋白质会失去其原来的功能，这是因为在高温下，多肽链不能折叠成正常的构象，因而失去活性。这称为温度敏感突变型(temperature-sensitive mutation)。显现突变型性状的温度称为限制性温度(restrictive temperature)，而保持原来正常性状的温度则称为许可温度(permissive temperature)，许多异常血红蛋白(abnormal hemoglobin)就是由错义突变产生的；另一方面，冷敏感(cold-sensitive)型突变则在较高温度下产生正常的蛋白，而在低温下则会产生失活的蛋白。

当发生无义突变时，便会改变生成的多肽长度，视无义突变发生的位置而异。如果在靠近起始不远处的有义密码子突变为无义密码子，则生成很短的片段；反之，如果是在靠近末端处，则生成的多肽链长度与正常的无大差异。如果无义密码子突变为有义密码子，则生成的多肽链比正常的长。

二、抑制基因突变

由突变产生的有害效果常可由第二次突变而使之恢复原来的性状，这也称为抑制或校正

(suppression)。这第二次突变可以是简单地把第一次突变所改变了的核苷酸顺序变回原来状态。但有些情况则较为复杂，它是在染色体的另一位点或另一个基因发生突变而消除了或抑制了第一次突变的效果，这称为抑制基因突变(suppressor mutation)。抑制基因突变可分为基因内抑制(intragenic suppression)和基因间抑制(intergenic suppression)二类型。两种类型的抑制均是使由于第一次突变所产生的失活蛋白质回复至原来的具活性(或部分活性)的蛋白质，但两者的机理则回然不同。

(一)基因内抑制

基因内抑制是在同一基因内发生第二次突变，从而抑制或校正第一次突变所产生的伤害，这种可被基因内抑制所回复的第一次突变，常是由插入或缺失单个核苷酸引起的。这便使在突变位点以后的密码子的解读发生移码(frameshift)，生成一段氨基酸顺序不同的多肽。而且，这样的移码常会产生新的无义密码子，结果使翻译至此中断而生成较短的不成熟的多肽链。如果在第一次突变位点附近发生第二次突变(插入或缺失核苷酸)，便有可能使第一次突变位点以后的密码子回复至正常，从而抑制了第一次突变(图 8-7)。从图 8-7 看出，由于第一次突变(突变型 DNA_1)缺失了一个核苷酸，使突变部位以后发生移码(密码子向后移 1 个碱基位)，因而，产生错义密码子和错误的氨基酸，也可能出现无义密码子而使多肽链过早终止，于是产生一条不完全的和无活性的多肽链。如果在缺失核苷酸附近发生第二次突变(抑制突变)，增加一个核苷酸，便会使以后的密码子恢复正常。虽然在两次突变点之间可能出现一些不正常的密码子，但由于密码子的简并性，很可能它们仍是编码同一氨基酸，如果是这样，便会产生具正常功能的多肽链。

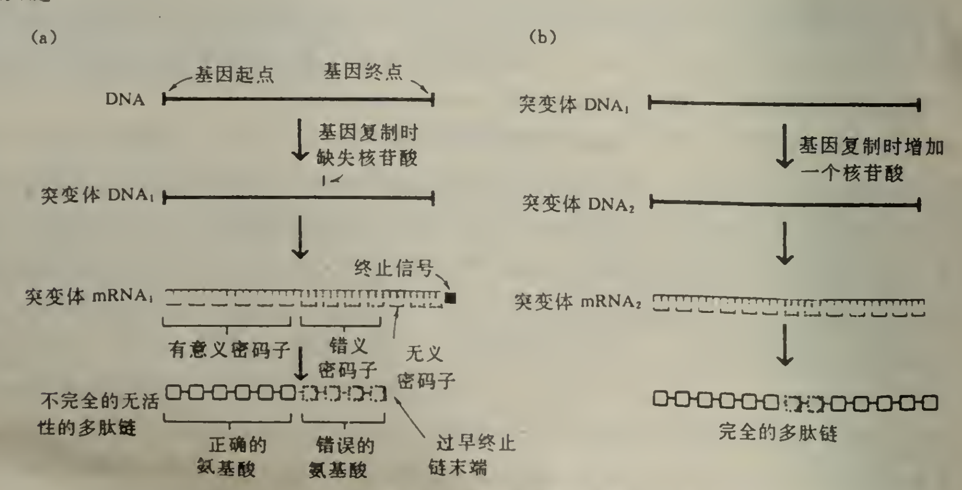


图 8-7　核苷酸缺失或插入突变的基因内抑制

(a)单个核苷酸缺失突变的结果；(b)第二次突变增加一个核苷酸抑制了第一次突变的伤害。其中两个氨基酸被置换，如果不是发生在要害部位，则多肽链将保持部分或完全的活性。

如果第一次突变是错义突变，那么也可由第二次错义突变而使之回复。这是因为第一次突变改变了密码子，使生成的多肽链中含一个错误的氨基酸，这便可能改变蛋白质的三维结构，从而使之失活。如果在同一基因内发生第二次突变，便可能使该蛋白的功能部位恢复原来构象，从而恢复其活性。例如，大肠杆菌的色氨酸合成酶由 A，B 两条链组成，这两条链单独存在便失去活性。A 链含 267 个氨基酸。当用诱变剂使它的第 211 个氨基酸即甘氨酸变为精氨酸

(密码子 GGA 变为 AGA)时,便失去活性。如果再发生突变使精氨酸变为丝氨酸(密码子 AGA 变为 AGU),便完全恢复活性;如果精氨酸变为苏氨酸(密码子 AGA 变为 ACA),也可部分恢复活性(图 8-8)。

(二)基因间抑制

基因间抑制比基因内抑制的情况更复杂,它是由于在另一基因发生突变而产生抑制的。这另一基因称为抑制基因(suppressor gene)。抑制基因的作用并不是改变第一次突变基因上的碱基顺序,而是由其它方式产生抑制的,有下列的不同类型:

1. 终止密码子的错读(misreading)　这是由于抑制基因改变了 mRNA 模板的解读方式而抑制突变的。3 个终止密码子都有抑制基因,这些抑制基因的作用是把终止密码子解读成某种氨基酸。已知 UAG 密码子有 3 个抑制基因:一个把 UAG 密码子解读为丝氨酸;另一个解读为谷氨酰胺;第三个解读为酪氨酸。这是因为抑制基因编码生成不正常的 tRNA,使携带这些氨基酸的 tRNA 的反密码子发生了变化。如 $tRNA^{Tyr}$ 基因发生突变,将反密码子由 3′-AUG-5′变为 3′-AUC-5′,这便使它可以识别 UAG 终止密码子,把它解读为酪氨酸(图 8-9)。同样,其它 2 个抑制基因也使 tRNA 的反密码子的一个碱基改变,使它们能将终止密码子 UAG 解读为丝氨酸或谷氨酰胺。

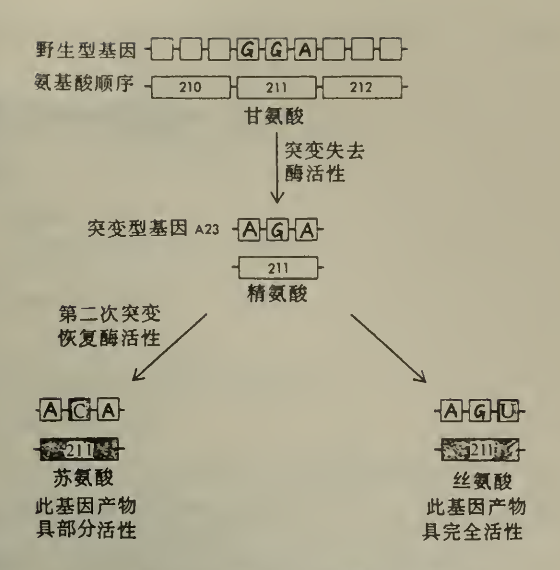


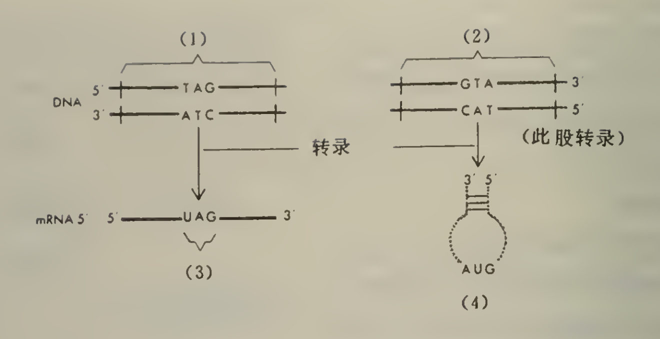
图 8-8　基因内抑制恢复色氨酸合成酶活性

终止密码子 UAA 也可以通过 tRNA 的突变而被抑制,例如,酪氨酸和赖氨酸的密码子和 UAA 只差一个碱基,通过 tRNA 的突变,可以将 UAA 解读成酪氨酸或赖氨酸。

终止密码子 UGA 也是通过突变 tRNA 分子被抑制的,这就是 $tRNA^{Trp}$,它使 UGA 解读成色氨酸。但这与上述的不同,不是改变了反密码子的碱基,而是把 tRNA 分子中第 24 位的 G 改变为 A。但这样的改变如何能影响其与密码子的配对还不清楚。从 $tRNA^{Trp}$ 的三维结构看,24 位是在 L 形结构的折角处,它能影响距离 1.5 nm 处的密码子-反密码子相互作用,说明核糖体上的 mRNA 的正确解读,要求 tRNA 分子有微小的构象变化。此外,即使正常的 $tRNA^{Trp}$ 也能解读 UGA 终止密码子,但其频率很低。因此,终止密码子 UGA 叫做"渗漏"(leaky)密码子。

既然具有无义抑制基因的细胞含有发生了突变的 tRNA,那么,与这些 tRNA 对应的密码子如何能正常地被解读呢?目前,对酪氨酸 UAG 抑制基因已有所了解,发现编码 $tRNA^{Tyr}$ 的基因有 3 个,其中一个编码主要的 $tRNA^{Tyr}$,其它两个是重复基因(duplicate gene),只编码生成少量的 $tRNA^{Tyr}$。在大肠杆菌的染色体中,这两个重复基因距离 200 bp。抑制突变就是发生在两个重复基因的一个上面。UGA 的抑制则与此不同,因为起抑制作用的 $tRNA^{Trp}$ 仍能解读色氨酸密码子 UGG。

无义抑制的作用,可视为抑制基因 tRNA 与蛋白质合成的释放因子之间的竞争作用。当终止信号进入核糖体的 A 位点后,它可能与抑制基因 tRNA 结合而解读成氨基酸,这时肽链便继续延长;它也可能与释放因子结合而终止肽链的延长。这便发生竞争,视何者先到达而结



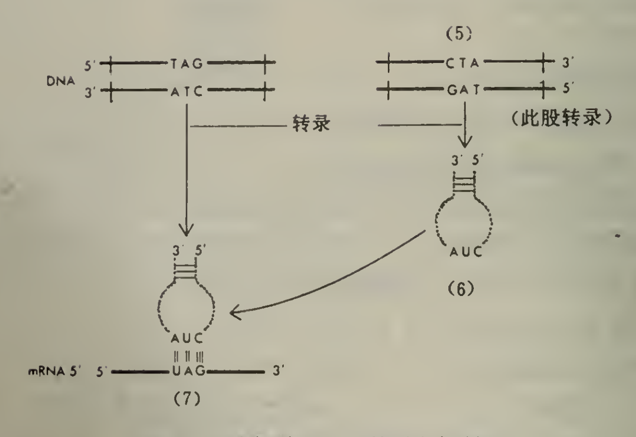


图 8-9　酪氨酰 tRNA 的无义抑制

(1)基因突变使酪氨酸密码子变为无义密码子;(2)编码酪氨酰 tRNA 的基因;(3)突变 mRNA 含无义密码子;(4)酪氨酰 tRNA;(5)编码反密码子区发生突变;(6)突变产生的酪氨酰 tRNA;(7)无义密码子被抑制,解读为酪氨酸。

果不同。UAG 和 UGA 密码子的抑制效率是相当高的,在有抑制基因 tRNA 存在时,约有 50%的 UAG 和 UGA 被解读为氨基酸。但 UAA 的抑制通常只有 1%～5%。起先人们以为,由于 UAG 和 UGA 的抑制效率很高,它们很少起终止密码子的作用。因为如果它们常被使用,则抑制基因的存在将使肽链超常延长,细胞生长也就停止。后来人们发现,在许多基因的末端常常出现 UAG 和 UGA 密码子,虽然有它们的抑制基因存在,却不生成太多不正常的过长的多肽链。其原因尚不清楚。

另一个还没有在分子水平上解释清楚的问题是,终止密码子周围的核苷酸顺序如何影响其抑制效率。视无义密码子在 mRNA 上的位置的不同,其相应的抑制基因 tRNA 抑制的程度相差达 10 倍之多。而且,这种对终止密码子解读的"上下文效应"(context effect)并不限于含抑制基因的细胞。已知道一些正常的终止密码子也能被解读成氨基酸而使肽链延长,虽然其程度很低,但亦有生物学上的重要性。一个已充分证明的例子是噬菌体 $Q\beta$ 的 RNA 基因组翻译。这种噬菌体有两种外壳蛋白,一种主要的(分子量约 14 kD)和一种次要的(分子量约 38 kD),它们的 N-末端氨基酸顺序均相同。这种次要的外壳蛋白是由主要外壳蛋白的 UGA 终止密码子偶然被解读为色氨酸,遂使肽链延长成为次要外壳蛋白(图 8-10)。这种次要蛋白是病毒侵

染所必需的。图 8-10 表明，约有 5%的 UGA 被解读为色氨酸，使肽链延长，生成含 327 个氨基酸的次要外壳蛋白(A 蛋白)。

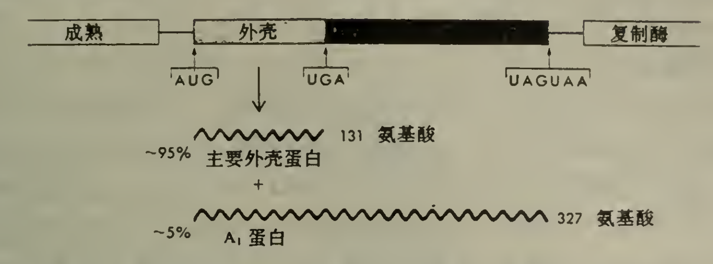


图 8-10 图示噬菌体 $Q\beta$ 基因组中主要外壳蛋白的终止密码子 UGA 被抑制生成次要外壳蛋白

正常的终止密码子本身也可能发生突变而使之成为可被解读成某种氨基酸的密码子。由于 mRNA 的 3′端常含有不翻译的区段，这样的突变便使多肽链不能正常终止，而生成一条特别长的多肽链。例如，人类的血红蛋白的 α 链含 141 个氨基酸，其基因的终止密码子是 UAA。但当 U 变为 C 时，它便变为编码谷氨酰胺的密码子 CAA，肽链的第 142 位便成为谷氨酰胺，并继续延长至 172 个氨基酸。

表 8-5 大肠杆菌的甘氨酰-tRNA

基因	tRNA 产物	反密码子	被识别的甘氨酸密码子	产生的错义抑制基因	tRNA 产物的反密码子	被识别的精氨酸密码子
glyU	$tRNA_1^{Gly}$	3′-CCC-5′	GGG	*gly Usu*	3′-UCC-5′	AGG
glyT	$tRNA_2^{Gly}$	3′-CCU-5′	GGA	*gly Tsu*	3′-UCU-5′	AGA
			GGG			AGG
glyV	$tRNA_3^{Gly}$	3′-CCA-5′	GGU			
			GGC			

2. 错义突变的抑制　错义突变亦可由 tRNA 突变而被抑制。例如，色氨酸合成酶 A 链的甘氨酸改变为精氨酸，此酶便失去活性。可由抑制基因突变而使之变回甘氨酸，并恢复酶活性。这也是由 tRNA 突变产生的。在大肠杆菌内有 3 种 tRNA^{Gly}，其反密码子均不同，分别识别不同的密码子(表 8-5)。从表 8-5 看出，$tRNA_1^{Gly}$ 识别密码子 GGG，但 GGG 也可被 $tRNA_2^{Gly}$ 识别。所以，对蛋白质合成来说，$tRNA_1^{Gly}$ 并不是必需的。当它的反密码子 3′-CCC-5′改变为3′-UCC-5′时，便可以解读精氨酸的密码子 AGG，而不会影响对甘氨酸密码子 GGG 的解读。

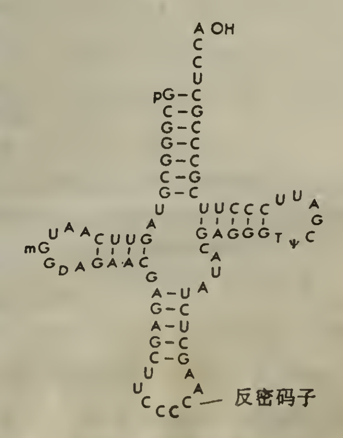


图8-11 移码抑制基因 $tRNA_{SufD}^{Gly}$的三叶草形结构和核苷酸顺序

另一些例子是由于 tRNA 的突变，使它携带错误的氨基酸而产生抑制。在 tRNA 分子的不同部位均可发生改变，使它被错误的氨酰-tRNA 合成酶识别，从而运载了错误的氨基酸。

3. 移码抑制　mRNA 上的密码子是以 3 个核苷酸为一组，组成一个三联体，称为码组或框架(frame)。如果发生突变，核苷酸增加了 1 个或 2 个，那么便会使以后的密码子码组发生改变，称为移码(frameshift)。例如，正常 mRNA 为CAU⋮CAU⋮CAU⋮……，如果在 U 之后插入一个 U，三联体便变为 CAU⋮UCA⋮UCA⋮UXX……，即密码子向后移了一个位置。缺失核苷酸也同样会发生移码。但如果插入或缺失的核苷酸是 3 个或其倍数，则不会发生移码。

由插入核苷酸而产生的移码突变，也可由抑制基因使之消除。在这里，也是由 tRNA 的突变起作用。例如，移码甘氨酰-tRNA(称为 $tRNA^{Gly}_{SufD}$或 $tRNA^{Gly}_{GGGG}$)。它的反密码子多了一个碱基，由原来的 CCC 变为 CCCC(图 8-11)。它是由基因 *glyU*(见表 8-5)发生插入突变产生的。由于 $tRNA^{Gly}_{GGGG}$的反密码子有 4 个碱基，它便可以和 mRNA 上发生插入突变的 4 个碱基配对，在核糖体上由 A 位点移至 P 位点时，一次移动 4 个碱基。这样，以后的密码子便恢复正常。脯氨酰-tRNA 也有类似情况。

4. 由核糖体错读产生的抑制　核糖体的结构与遗传密码的正确解读有密切关系。如果由于突变引起核糖体 30 S 亚基的蛋白质的氨基酸顺序发生变化，便会影响它的正常结构，从而导致与错误的氨酰-tRNA 分子相结合。一个明显的例证是抗菌素链霉素的作用。将链霉素加入到活细胞或体外系统中，便会使遗传密码的翻译发生错误，产生无活性的蛋白质，导致细胞死亡。链霉素与核糖体的 30 S 亚基结合，干扰了其正常的 mRNA-tRNA-核糖体相互作用，从而导致错读。

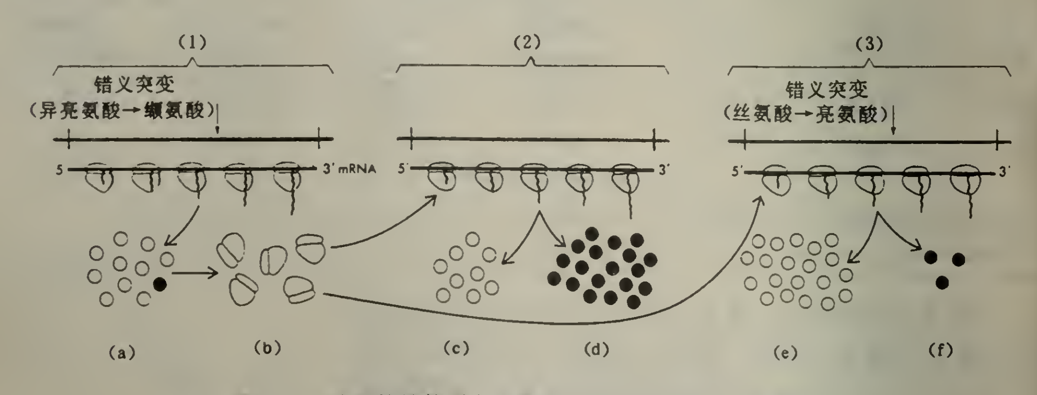


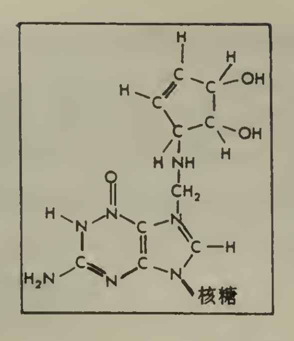
图 8-12　编码核糖体蛋白的基因发生错义突变而引致抑制

(1)突变体基因 *A* 编码一种核糖体蛋白的氨基酸；(2)基因 *R* 编码具活性的 R 酶的氨基酸；(3)突变体基因 *X* 编码无活性的 X 酶的氨基酸。

(a)突变体核糖体蛋白；(b)突变体核糖体可偶尔错读遗传密码，将亮氨酸解读为丝氨酸；(c)由于亮氨酸→丝氨酸产生的少量无活性的酶；(d)大量的正常具活性的酶；(e)大量的无活性的酶；(f)由于亮氨酸→丝氨酸产生少量的具活性的酶。

图 8-12 表示出，由于编码核糖体蛋白的基因 *A* 发生错义突变，使异亮氨酸改变为缬氨酸，产生核糖体蛋白突变体。这个核糖体突变体偶然被引起错读，使原来应解读为亮氨酸的密码子解读成丝氨酸(图 8-12 左)。这便会使小部分具活性的 R 酶由于核糖体错读发生亮氨酸→丝氨酸的改变而失去活性，但大部分酶仍是正常的(图 8-12 中)。如果另一编码 X 酶的基因 *X* 发生错义突变，使丝氨酸改变为亮氨酸，因而产生无活性的酶，由于核糖体突变体也可将亮氨酸变为丝氨酸，从而恢复酶活性。但这个抑制很少发生，所以，只生成很少量的具活性的酶。

5. 抑制基因引起正常基因的错读　抑制基因的产物并不选择性地对突变体基因的mRNA模板起作用，它们也能对野生型 mRNA 模板的解读起作用，因而干扰正常蛋白质的合成。一般来说，这样的变化对正生长着的细胞为害不大，因为通常生成正常蛋白质的量大大多于不正常的量。但是，即使是生成小量不正常蛋白质，对细胞也是没有什么好处的。



关于抑制基因的功能可归纳如下：一个基因由突变而成为抑制基因。在这种突变发生之前，这个基因是正常的和具活性的。它可以编码一种特异的 tRNA，或一种核糖体蛋白，或一种氨酰-tRNA 合成酶，也可能包括修饰 tRNA 的酶。现在已证明，在反密码子中的或其附近的碱基被修饰，均可影响抑制(图 8-13)。图 8-13 为果蝇(*Drosophila*)的 $tRNA^{Tyr}$，当它的反密码子的第一位为 G 时，它偶然可以解读终止密码子 UAG，从而产生抑制；当用 Q(*Queuosine*)代替 G 时，便失去抑制活性。生物在进化过程中形成所有这些基因，使得其产物具有最佳的构型，保证遗传密码能正确解读。如果发生一次突变，产生改变了的产物，使错读增多，这时这些基因便转变为抑制基因，使由于第一次突变产生的伤害得以校正。

图 8-13　果蝇 $tRNA^{Tyr}$ 的结构

第六节　遗传密码的改变

上文在谈到遗传密码的破译时指出，在用体外合成蛋白质系统研究密码子时，不管所用的无细胞提取液是来源于细菌或高等生物，加入多聚 U 总是得出多聚苯丙氨酸肽链，加入多聚 C 则总是得出多聚脯氨酸，加入多聚 A 得出多聚赖氨酸。这似乎意味着遗传密码在目前存在的从低等到高等的各种生物中都是通用的。

但是，Barrell 等在 1979 年发现在人的线粒体中，通用的密码子 AUA 却是编码甲硫氨酸而不是异亮氨酸，UGA 不是作为终止密码子而是编码色氨酸。以后，许多的研究发现在多种生物中有些密码子与表 8-3 列出的通用密码不同，这些非通用的密码子在多种生物体中均有发现，而且主要是在线粒体的遗传密码中发现的。这些遗传密码的自然变化可分为 2 个类型：一个类型是终止密码子 UGA，UAA 和 UAG 改变为编码氨基酸的密码子。这在多种生物的线粒体遗传系统、一些原生动物的核质系统、以及一些原核生物的遗传系统中均有发现；另一类型是改变有义密码子的编码意义，即编码的氨基酸与表 8-3 列出的不同，它存在于线粒体的遗传系统中。

一、线粒体的遗传系统

线粒体 DNA 中改变的密码子如表 8-6 所示。从表中看出，许多种生物的终止密码子

UGA 改变为编码 Trp，而在牛、鼠等动物中的 AGA 和 AGG 则改变为终止密码子。AUA 在一些动物和酵母中作为 Met 的密码子，AUA，AUU 和 AUC 也作为鼠线粒体基因的起始密码子。在高等植物线粒体内，则用另一密码子变异体来编码 Trp，例如，玉米的线粒体中的细胞色素氧化酶亚基Ⅱ的基因中即含有 3 个 CGG 密码子用以编码 Trp。一些对莱因衣藻线粒体基因的研究未发现非通用的密码子，亦未发现 UGA 和 CGG 密码子(表 8-6 中的"?"号)，很可能它的终止密码子也发生变异。

表 8-6　线粒体中的密码子变异体

生物体	密码子				
	UGA	AUA	AGA AGG	CUN	CGG
通用密码动物	终止	Ile	Arg	Leu	Arg
脊椎动物	Trp	Met	终止	+	+
果蝇	Trp	Met	Ser	+	+
酵母					
Saccharomyces cerevisiae	Trp	Met	+	Thr	+
Torulopsis glabrata	Trp	Met	+	Thr	?
Schizosaccharomyces pombe	Trp	+	+	+	+
丝状真菌	Trp	+	+	+	+
锥虫	Trp	+	+	+	+
高等植物	+	+	+	+	Trp
莱因衣藻	?	+	+	+	?

注：(1)"+"号表示与通用密码子相同；(2)N 表示任一种核苷酸；(3)"?"表示在这种线粒体基因组中未观察到这个密码子。

按照变偶假说，至少要有 32 种 tRNA 才能与 64 个密码子配对。但在脊椎动物线粒体中只发现有 22 种 tRNA，因此，便有一套不同的变偶规律，使 22 种 tRNA 与 64 个密码子配对。在非混合型密码子族(密码子头 2 个碱基相同，第三个不同)中，只用一个 tRNA 与之配对，其反密码子的第一个(变偶)碱基为 U，这 U 可能与 4 个密码子第三个碱基配对，也可能根本不配对。另一些 tRNA 识别第三个碱基为 A 或 G 的密码子，或为 U 或 C 的密码子。这样，所有 tRNA 可以识别 2 或 4 个密码子。由于线粒体的 tRNA 可以识别 2 个密码子，在通用密码中的终止密码子 UGA 便改变为编码 Trp，Ile 的密码子 AUA 改变为编码 Met。

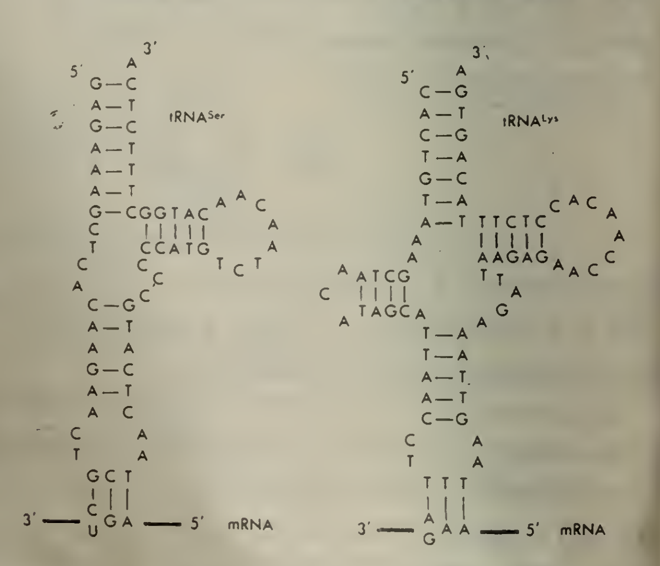


图 8-14　人类线粒体的 2 种 tRNA 的基因核苷酸顺序以三叶草形表示(转录后在 3′端加上 CCA)。

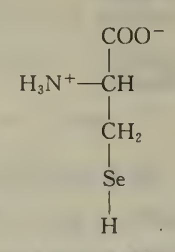
线粒体 tRNA 的结构也与非线粒体的不同。例如，常缺少 GTΨCRA 顺序。D 环和 TΨC 环中的一些核苷酸也发生变化。在非线粒体 tRNA 的三叶草形结构中，TΨC 环总是含 7 个核

苷酸，但线粒体 tRNA 可含 3～9 个。还有线粒体的一种丝氨酸 tRNA 完全缺少 D 臂（图 8-14）。这表明线粒体 tRNA 结构及其与核糖体的相互作用方式，均与非线粒体的不同。

二、非线粒体的遗传系统

除线粒体外，在非线粒体的遗传系统中也发现有密码的改变。例如，原核生物 *Mycoplasma capricolum* 的 UGA 密码子变为编码 Trp。在纤毛原生动物嗜热四膜虫（*Tetrahymena thermophila*）中，UAA 和 UAG 不作为终止密码子而改变为编码 Gln。

三、编码的位点专一性变异



硒代半胱氨酸

在一些遗传系统中观察到在特定位点的密码变异，包括用无义密码子来编码氨基酸，以及非三联体的翻译。例如，在大肠杆菌中的甲酸脱氢酶是一种硒蛋白，含有硒代半胱氨酸（selenocysteine Se-Cys）。它是由基因 *fdhF* 编码的，其结构基因含 715 密码子。试验证明其肽链的第 140 个氨基酸是由 UGA 编码的 Se-Cys。在哺乳动物中也发现 UGA 编码 Se-Cys。另一例子是大肠杆菌的释放因子 2（RF-2，见第九章）的基因，它的第 26 个密码子是 UGA（原是终止密码子），但在这个密码子发生移码，解读为 4 个碱基的 UGAC，编码为 Asp（Asp 的通用密码子是 GAC）。以后即按这个新的（+1）框架进行翻译。

一般来说，在绝大多数情况下各种生物仍是用通用的标准密码，只在少数情况下出现密码子的变异。这说明地球上的生物在遗传上是同源的。

第七节　重叠基因与重叠密码

由于 DNA 的密码子是连续的，在密码子之间没有间断，这样，一条 DNA 链便可以有 3 个解读框架，编码出 3 条不同的多肽链。但大多数的 DNA 只用其中一个框架，编码出一条多肽链。而且，一般说来各个基因也是不重叠的。但是这个一般规律有时也有例外。在一些病毒中发现同一 DNA 碱基顺序可以编码出两条不同的多肽链。这是由于有些基因可以完全埋藏在另一个基因之内。这表明一段 DNA 顺序有多个功能。例如，1976 年 Barrell 等发现，在噬菌体 *ΦX*174 环型的单链 DNA 中，*E* 基因的 237 个核苷酸完全包含在含有 456 个核苷酸的 *D* 基因之内，但它们的解读框架不同。1977 年 Sanger 又发现 *B* 基因的 260 个核苷酸完全位于含 1 546 个核苷酸的 *A* 基因之内。此外又发现另一个 *K* 基因则跨越在 *A* 基因和 *C* 基因之间（图 8-15）。基因 *E* 与基因 *D* 共用一段相同的碱基顺序，但解读框架不同。图 8-16 表示这 2 个基因的部分碱基顺序中的密码子及其编码的氨基酸。从图中可以看出，*E* 基因

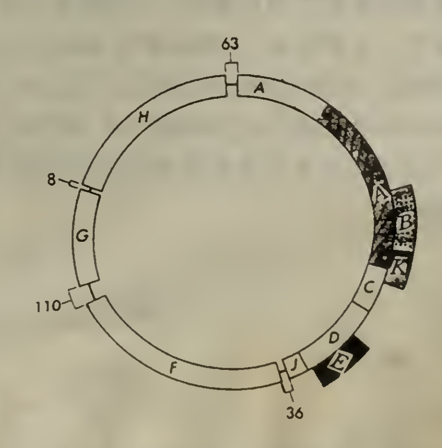


图 8-15　噬菌体 *ΦX*174 的环型 DNA
B 基因在 *A* 基因序列内，*E* 基因在 *D* 基因序列内，但均用不同的解读框架。*K* 基因跨越在 *A* 和 *C* 基因之间。

的解读框架向 3′ 端移动了一个碱基。以后在多种病毒 DNA 中也发现有类似情况，如噬菌体 λ、猿猴病毒 40(SV40)、以及 RNA 噬菌体 $Q\beta$ 和 $Q17$、噬菌体 GA 等。有人认为上述这种一个基因内包含另一个基因的情况只在病毒内存在，因为在病毒外壳蛋白所包围的内部空间很小，必须充分利用其所包含的有限 DNA 以编码必需的多种蛋白。但是，在编码多于一种蛋白质的细菌 mRNA 中，也常发现在一小区段内有几个核苷酸的重叠，在其中一个编码区终止，而另一个起始，如图 8-17 所示。

D 基因的 --- Val — Glu — Ala — Cys — Val — Tyr — Gly — Thr — Leu — Asp — Phe ---
解读框架 5′---GUU GAG GCU UGC GUU UAU GGU ACG CUG GAC UUU G---3′

E 基因的 ---G U U G A G G C U U G C G U U U AUG GUA CGC UGG ACU UUG---
解读框架 Met — Val — Arg — Trp — Thr — Leu ---

图 8-16　*ΦX* 174 DNA 的基因 *D* 和基因 *E* 的 mRNA 转录本部分核苷酸顺序的解读框架

重叠基因是怎样翻译的呢？是否在同一条 mRNA 链上由核糖体按 2 个不同框架解读同一个基因，还是在不同 mRNA 上分别解读不同基因？对噬菌体 MS2 的重叠基因的研究表明，重叠基因开始进行翻译之前，"母"多肽的合成必须由错读导致终止。噬菌体 MS2 的 RNA 含 3

trpE-trpD　　CAGGAGACTTTC T GATGGCT

trpB-trpA　　CGAGGGGAAATCTGATGGAA

图 8-17　大肠杆菌色氨酸操纵子中几个顺反子间连接处的基因重叠

起始及终止密码子分别以 ┌─┐ 及 └─┘ 号表示。

个主要基因，被顺反子间的间隔区隔开(图 8-18)。另一重叠基因，是在第二个基因(外壳蛋白基因)的末端开始，跨过隔离区，插入第三个基因(复制酶基因)内。这重叠基因是编码一种引起寄主细胞溶胞的蛋白质的。这溶胞蛋白基因相对于外壳蛋白基因和复制酶基因是 +1 码组(即向前移一个核苷酸)。溶胞蛋白基因的表达，必须与外壳蛋白的产生偶联。当翻译外壳蛋白的核糖体进行至距溶胞蛋白基因不远处时，便发生移位错读(移码)，结果遇上一个无义密码子，多肽链的合成便终止并释放出来，然后在 AUG 密码子指导下开始溶胞蛋白的合成。但目前还不知道是否所有的重叠基因的解读均要发生移码错误，也有可能核糖体在重叠基因起始处重新结合。

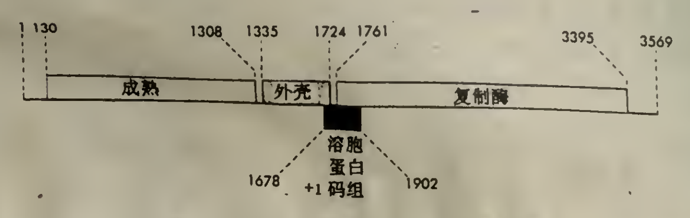


图 8-18　RNA 噬菌体 MS 2 的遗传图

一些小基因组的解读有时亦利用核糖体移码，使得由同一 DNA 顺序可以编码两条不同的多肽链，这便增加其编码的效能。例如，噬菌体 T 7 有一条含约 30 个基因的线状双链 DNA 基因组，由同一基因可产生一种主要的和一种次要的外壳蛋白。这两种蛋白的 N 端氨基酸顺

序相同，但C端的顺序和长度均不同。次要的外壳蛋白较长，这是由于在编码主要外壳蛋白区段的近末端处有约10%的核糖体移入－1码组(向后移一个核苷酸)，使肽链超越终止密码子而继续延长约20个氨基酸，至遇到另一个终止密码子才停止。所以，次要外壳蛋白比主要蛋白长约20个氨基酸。

参考文献

1. 徐铃，彭朝晖编. 基因的分子生物学. 第四章. 北京：中国科学技术出版社，1992.
2. Fox, T. D. Natural Variation in the Genetic Code. Ann. Rev. Genet. 1987. 21:67～91.
3. Lehninger, A. L. et al. Principles of Biochemistry. (2nd. ed.). Chapters26. Worth Publishers, New York: 1993.
4. Watson, J. et al. Recombinant DNA. (2nd. ed.). Chapters3. 1992.
5. Watson, J. A. et al. Molecular Biology of the Gene. (4th ed.). Chapters15. The Benjamin/Cummings Pub. ,Co. Inc. California: 1987.

（李明启）

第九章　RNA 的翻译——蛋白质的生物合成

第一节　引　　言

蛋白质是生命活动的重要物质基础，要不断地进行代谢和更新，因此，细胞内利用 20 种氨基酸进行蛋白质合成便成了生命现象的主要内容。细胞内每个蛋白质分子的生物合成都受细胞内 DNA 的指导，但是贮存遗传信息的 DNA 并非蛋白质合成的直接模板。它是经转录作用把遗传信息传递到信使核糖核酸的结构中，然后再经翻译作用将遗传信息从信使核糖核酸传递到蛋白质结构中去，使合成的产物具有一定的正确无误的结构。在遗传信息传递的过程中，包括了如何将 DNA 分子中的碱基顺序转变为多肽链中的氨基酸顺序的信息传递问题，还包含了蛋白质生物合成的实际过程：即合成的起始；将氨基酸按一定的顺序连接；肽链的终止；完整的肽链从合成装置上的释放；肽链的折叠以及新合成肽链的修饰等等的被称之为翻译的问题。在整套遗传信息传递过程中，遗传信息传递问题已由前文讨论过了。本章则讨论核糖核酸的翻译问题，即蛋白质的生物合成。

本世纪 50 年代初期，Caspersson 等人就发现处于生长状态的组织和分泌蛋白质的细胞，蛋白质合成的速度极高，同时发现 RNA 的含量也特别丰富。这些现象使学者们得知，蛋白质的生物合成与核酸有着密切的关系。经过近 40 年的努力，对于这个问题的认识已比较清楚。现在知道蛋白质生物合成过程要比“复制”与“转录”复杂得多，需要大约 200 多种生物大分子，其中包括有核糖体（由 RNA 和蛋白质组成）、信使 RNA（messenger ribonucleic acid，mRNA）、转移 RNA（transfer ribonucleic acid，tRNA）、氨酰-tRNA 合成酶（aminoacy1-tRNA-synthetase）以及一些辅助因子，即可溶性蛋白质分子（起始因子）（initiation factor，IF）、延伸因子（elongation factor，EF）、释放因子（终止因子）（release factor，RF）等参加的协同作用。

蛋白质合成的场所是在核糖体，蛋白质合成的原料是细胞中的 20 种氨基酸，反应所需的能量由 ATP 与 GTP 提供。

蛋白质生物合成的早期研究工作都是用大肠杆菌的无细胞体系（cell-free-system）进行的，所以，对大肠杆菌的蛋白质合成机理了解最多，真核生物的蛋白质合成的机理与大肠杆菌有许多相似之处，但也有不少差异，下面即详细讨论原核生物与真核生物的蛋白质的生物合成。

第二节　核糖体及核糖体核糖核酸结构

一、核糖体与多核糖体

核糖体是由几十种蛋白质和几种 RNA 组成的一种亚细胞颗粒。它是蛋白质生物合成的场所。基本上不含脂肪。研究得最详细的是大肠杆菌的核糖体。一个迅速生长的大肠杆菌细

胞内有约 15 000 个核糖体，每个核糖体的分子量略小于 3×10^3 kD，故核糖体大约共占细菌细胞总质量的 1/4。所以，在总的细胞合成物质中有相当大部分是用于制造核糖体。在单个核糖体上每次只能合成一条多肽链。在最适条件下，合成一条含 400 个氨基酸的多肽链（分子量约为 40 kD）约需 10 s。合成好了的多肽链便被释放，游离的核糖体立即用于另一轮蛋白质的合成。

利用超速离心与其它分离技术对核糖体组分的化学结构已经基本弄清，从形状看真核生物细胞质的核糖体与原核生物的核糖体一样都是由大小 2 个亚基组成。大亚基的大小约为小亚基的 2 倍（图 9-1）。2 个亚基均含有 RNA 和蛋白质。在大肠杆菌内，RNA 和蛋白质的比例

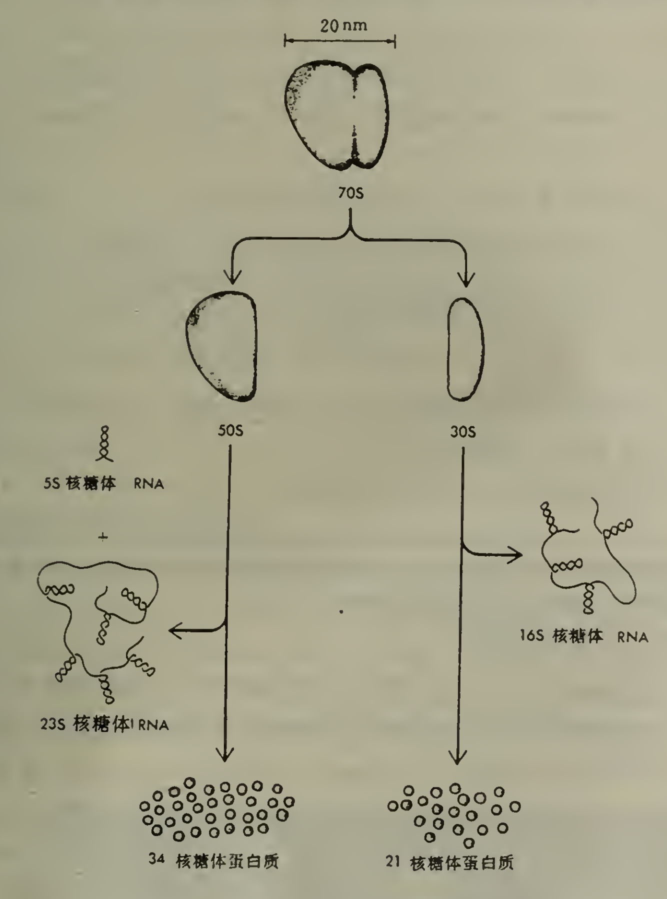


图 9-1 *E. coli*（大肠杆菌）核糖体结构

由于核糖体在离心时是以 70 S 沉降，故称其为 70 S 核糖体。S 代表 Sendberg，它是沉降速度的标准。同样，30 S 与 50 S 代表核糖体小亚基与大亚基的沉降系数。5 S，16 S 和 23 S，是小亚基与大亚基中 RNA 分子的沉降系数。5 S，16 S，23 S 核糖体 RNA 分子量分别为 40，500，1 000 kD；70 S，50 S，30 S 的分子量分别为 2 700，1 600，900 kD。所有细菌核糖体的大小与 *E. coli* 相似，具有 30 S 和 50 S 亚基。在高等生物中，核糖体稍大些（80 S），具有 40 S 与 60 S 亚基。

约为 2 : 1，在其它许多生物体中则为 1 : 1（表 9-1）。大小亚基均含有许多不同的蛋白质。小亚基（30 S）由 1 种 RNA（16 S，1 542 个核苷酸）和 21 种蛋白质组成，大亚基（50 S）由 2 种 RNA（23 S，2 904 个核苷酸和 5 S，120 个核苷酸）和 34 种蛋白质组成。由于核糖体中 3 种 RNA 的序列已经测出，核糖体及其亚基的分子量分别是：小亚基蛋白质共 350×10^3 kD，16 S rRNA 为 500×10^3 kD，两者之和为 850×10^3 kD；大亚基蛋白质为 460×10^3 kD，23 S rRNA 与 5 S rRNA 为 990×10^3 kD，共为 $1\,450\times10^3$ kD，故核糖体总分子量是 $2\,300\times10^3$ kD。大肠杆菌小

表 9-1　核糖体的性质与核糖体亚基

种　类	$S_{20,w}$	大小(nm)(最大的)	分子量(MD)*	RNA/蛋白质(W/W)	轴向比	rRNA 分子量(MD)	rRNA $S_{20,w}$
核糖体							
原核生物大肠杆菌	70	22.5±2.5	2.6～2.9	2:1			
真核生物							
细胞质	80	28.0±2.8	3.4～4.5	1:1			
叶绿体	70	22.5±2.5	2.5～3.3	1:1			
线粒体	55～77	—	3.2～4.5	1.1:1.3			
核糖体小亚基							
原核生物大肠杆菌	30	22.0±2.2	0.95	2:1	2:1	0.6	16 S
真核生物							
细胞质	36～41	25.0±2.5	1.4	1:1	2:1	0.7	18 S
叶绿体	28～35	22.0±2.2	1.2	1:1	2:1	0.6	16 S
线粒体	28～40	—	1.1～1.7	1:3	—	0.3～0.7	12～14 S
核糖体大亚基							
原核生物大肠杆菌	50	22.5±2.2	1.75	2:1	1:1	1.17	23 S
真核生物						0.04	5 S
细胞质	60	28.0±2.8	2.1～3.1	1:1	1:1	1.2～1.75	25～28 S
						0.05	5.8 S
						0.4	5 S
叶绿体	46～54	22.5±2.5	2.4	1:1	1:1	1.1	23 S
						0.04	5 S
						0.03[a]	4.5 S[a]
线粒体	39～60	—	1.65～2.5	1:3	—	0.5～1.5	16～25 S
						0.04[b]	5 S[b]

[a] 4.5 S rRNA 相当于在真细菌内的 23 S rRNA 的 3′-末端的 100 个核酸，在高等植物的叶绿体中作为一个分别的种类存在。

[b] 这种品种仅出现在植物线粒体中，在其它线粒体核糖体是不存在的。

—表示无资料可引用；* MD 为百万道尔顿。

亚基上的 21 种蛋白质分别以 S 1～S 21 表示，大亚基上的 34 种蛋白质分别以 L 1～L 34 表示。在字母 S 与 L 后面的数码则表示蛋白质在双向电泳系统中的迁移率。这样，从小亚基得到的移动最慢的蛋白质就是 S 1 了(图 9-2)。

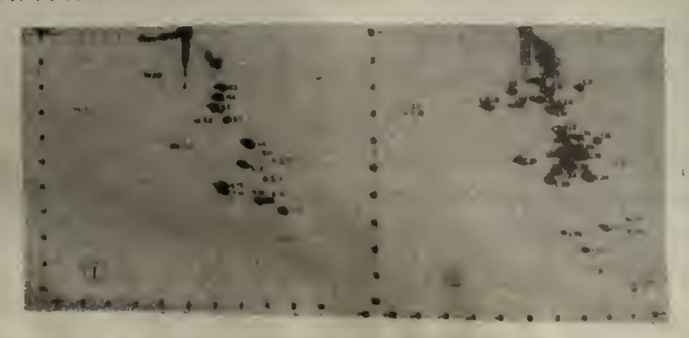


图 9-2　*E. coli* 核糖体小亚基(左)21 种蛋白质与大亚基(右)34 种蛋白质的双向凝胶电泳图谱

S 20 与 L 26 实际上为同一蛋白质，L 8 是 L 7/L 12 和 L 10 的复合物。一般来说，蛋白质的大小会随着数码增加而减小。这样，在 30 S 亚基上 S 1 是最大的蛋白质(分子量约为 60 kD)，S 21 是最小的蛋白质(分子量约为 8 kD)，而在 50 S 亚基上，L 1 是最大的蛋白质(分子量约为 25 kD)，L 34 是最小蛋白质(分子量约为 5 kD)。

在核糖体的大亚基中L 7 与L 12 的氨基酸数目与序列完全相同,只是L 7 的N 端为乙酰丝氨酸而L 12 的N 端是丝氨酸。蛋白质L 8 是L 7/L 12 和L 10 分子组成的复合物,这个复合物可能由于与核糖体内部的功能性结合有关,当分离核糖体时,这个复合物有时也不分开。此外,蛋白质L 26(每 5 个 50 S 颗粒才有 1 个L 26 分子)实际上是小亚基 30 S 颗粒上的蛋白质S 20,当 70 S 颗粒解离时,S 20 偶然离开 30 S 颗粒并且与 50 S 颗粒相结合。所以,在核糖体中,S 20=L 26。这样,大肠杆菌的核糖体的蛋白质共有 54 种。至于核糖体中蛋白质的数目,除L 7/L 12 具有 4 个拷贝外,每一种蛋白质只有 1 个拷贝。目前,54 种核糖体蛋白质全序列均已测出,得知小亚基大多数蛋白质是球状蛋白,带有 28%的 α 螺旋与 20%的 β 折叠。小亚基中,除S 1,S 2 与S 6 是酸性蛋白质外其它均为碱性蛋白质。而在大亚基中,只有L 7 与L 12 是酸性蛋白质,其它均为碱性蛋白质。现认为,荷负电荷的 RNA 与碱性蛋白质之间的相互作用有利于核糖体的稳定。

在大亚基上,现知L 7/L 12 具有特殊结构与功能,它是很多可溶性蛋白质因子出入大、小亚基间空腔的门户,包括起始因子、延伸因子和终止因子等都首先与L 7/L 12 相结合而发挥作用,并且在蛋白质合成的能量利用上起关键作用;L 16 被认为是与肽键形成有关。在小亚基上,S 1 能使 16 S rRNA3′端"发夹状"的二级结构散开,从而使核糖体核糖核酸(ribosomal RNA,rRNA)分子与 mRNA5′端的碱基配对处相结合。S 5 对 30 S 亚基的结构与功能也很重要,缺少S5 时,只能形成 30 S 亚基的前体 28 S 亚基,其活性仅为原来的 30%~50%,使蛋白质合成的起始受阻或减弱。S 5 与S 2 及S 4 的结合能提高蛋白质合成时延伸因子的 GTP 酶活性。S 4 与S 5 则与错读发生有关。S 12 可以降低蛋白质合成中的错读,因此,这 3 种蛋白质的相互作用,控制着蛋白质合成的精确性。

用氯化铯密度梯度离心,可以将核糖体蛋白质分为两部分,一种是核心蛋白(core protein,CP);另一种是脱落蛋白(split protein,SP)。将核心蛋白及脱落蛋白与RNA 放在一起保温,两者很快自行组装成为有功能的 30 S 与 50 S 亚基。脱落蛋白再用层析法又可将其分为酸性部分与碱性部分。这两种性质不同的部分在功能上是不同的。实验表明在小亚基中的酸性蛋白在蛋白质合成中仅起一种激活作用,而碱性蛋白在合成中则是重要的。与此相反,在大亚基上则酸性蛋白是蛋白质合成的重要组分,而碱性蛋白只起激活作用。

与原核生物不同,真核生物中细胞质与细胞器(主要是叶绿体和线粒体)的核糖体是不同的。这种不同反映在核糖体的总体大小与RNA 与蛋白质的组成上。高等真核生物细胞质的核糖体比细菌的大;蛋白质和 rRNA 的总含量比较高,大的 rRNA 的分子也比较长;蛋白质的种类数也比较多。表 9-1 列出了真核生物细胞质核糖体组分与大小的数据。核糖体小亚基(40 S)有 1 个 18 S rRNA 与大约 33 个蛋白质;大亚基(60 S)则由 1 个 28 S rRNA 和 1 个5.8 S rRNA 与 1 个 5 S rRNA,以及大约 49 个蛋白质组成。从表 9-1 可以看出,原核生物核糖体颗粒分子量的大部分是由 rRNA 提供的,在真核生物中,rRNA 也是核糖体颗粒分子量中的主要成分。

真核生物细胞器的核糖体则与细胞质的核糖体有所不同。比如叶绿体的核糖体与细菌核糖体大小相近,但它们的 RNA 比例要比细菌的大一些,而植物线粒体中的核糖体则比它周围的细胞质的核糖体略小(表 9-1)。低等真核生物(如真菌)核糖体总的大小还是比细菌的大,但是,哺乳动物的线粒体或两栖类动物的线粒体显得更小,总共只有 60 S,其中 RNA 的分子量比也较小。

在真核生物的细胞质中，核糖体常常与细胞骨架——一种纤维状的基质结合在一起。在有些真核细胞中，核糖体与内质网膜结合在一起。尽管核糖体在细胞内的存在方式可以不同，但其共同特点是，承担着蛋白质合成任务的核糖体在细胞中不是自由存在的，而总是直接或间接地与细胞结构结合在一起。

无论是原核生物还是真核生物，从细胞中可分离出 3 种类型的核糖体：核糖体、核糖体亚基和多核糖体(polysome)。细胞内蛋白质的生物合成正是通过这些核糖体循环进行的。在细胞质中，大多数核糖体处于非活性的稳定状态单独存在。只有少数与 mRNA 一起形成多核糖体。在多核糖体中，其大小变化，一般视 RNA 链的长短及核糖体的组装紧密程度而异，后者显然与核糖体在一个特定的基因的开端起始的频度有关，而这又随核糖体的结合位点不同而异。一般一条 mRNA 的最大利用率是每 80 个核苷酸有 1 个核糖体(图 9-3)。

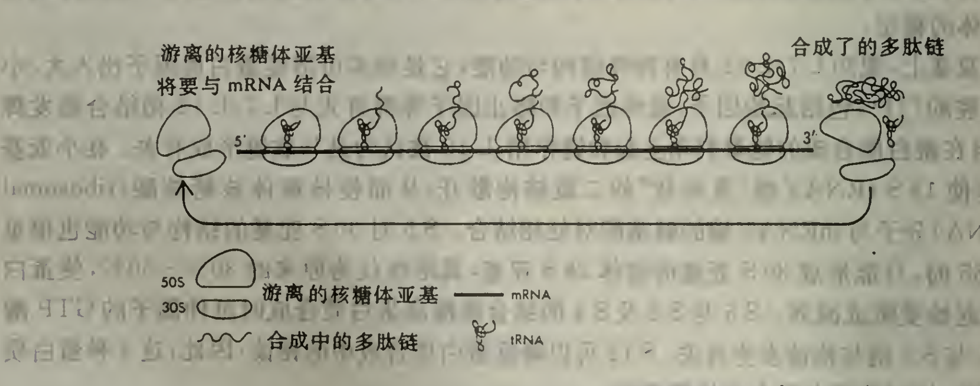


图 9-3 蛋白质合成时多核糖体的图示

mRNA 分子由右向左移动。

利用两种不同的电子显微镜对核糖体进行了观察。一种是对含有分离出的 70 S 核糖体、50 S 亚基和 30 S 亚基的视野做直接的观察(图 9-4)；另一种是用数学重组法(mathematical reconstruction methods)以解释天然存在的核糖体微晶体产生的图像。现在认为 50 S 亚基像一个半球形，从其平面伸出 3 个突起，30 S 比较扁平，可分为上 1/3 和下 2/3，在其结构的下部一侧伸出一个平台。图 9-4 和图 9-5 表示推导出的两个亚基的形状以及它们如何装配成 70 S 核糖体。

另一方面，利用重组技术、化学交联和化学修饰技术、荧光标记技术、重氢标记中子衍射技术、免疫电镜以及亲和层析等方法，确定了核糖体蛋白质彼此相邻情况。最近，Stark 等人用成角重组(angular reconstitution)与低温电镜(cryoelectron microscopy)结合方法，对大肠杆菌核糖体进行研究，得出一个 0. 3 nm 的高分辨率的立体图像(图 9-6)。这个图显示出核糖体呈直径约 2. 0 nm 的分枝状结构，在 50 S 亚基内广泛存在网状通道，它们是新生肽的出口处，Stark 称之为出口通道复合物(exit channel complex，ECC)(图 9-6)。有两个可能是进入 ECC 的位点，一个是在 L 1 突起与中央突起(CP)之间的裂隙；另一个则与第一个相邻，刚好在 CP 的下方。在 70 S 颗粒的外表面，ECC 具有若干个开口(图 9-6 b 中较大的开口是①～④)，在 50 S 颗粒左侧较低位置的 1 号开口是用免疫电镜观察到的新生肽链伸出处。

对于 16 S rRNA 在核糖体 30 S 亚基的装配如图 9-7 所示。图中显示出 30 S 亚基的“头”部、“颈”部、“躯干”、“足”部、“嘴”、“趾”以及“连接”等各部分。rRNA 的各螺旋用“h”表示，“h”后面的数字为螺旋的编号，从图 9-7 中可看出各螺旋在 30 S 亚基上的位置。

a

上部
1/3
裂口
平台
下部
1/3
中心突起部
A
小亚基 (S)
大亚基 (L)
70S 核糖体

上部
1/3
平台
凹部
中心突起部
B
b

图 9-4　用电子显微镜得到的核糖体形状的直接显影

a 是大肠杆菌核糖体的视野，它包括了在模型 b 中 A 或是 B 的取向，一些核糖体亚基也在视野中显示出来，用 S 代表小亚基，L 代表大亚基，视野是用电镜负染得到的，每一个核糖体是由一种重金属盐（铀盐）的轮廓确定的。

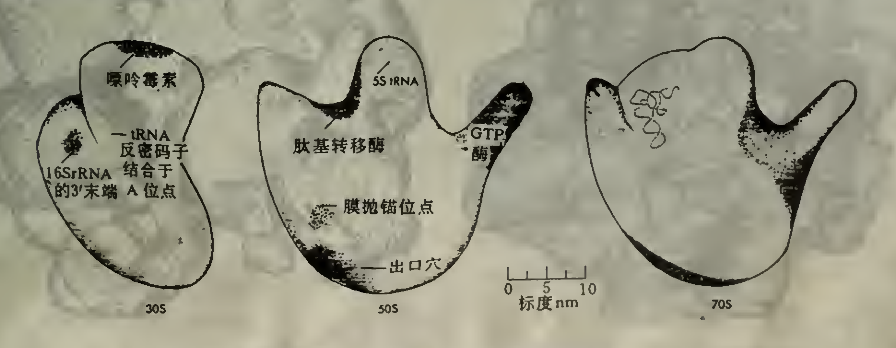


图 9-5　*E. coli* 核糖体 2 个亚基的形状

通过结合的方法将其功能位点定位，70 S 核糖体图示表示出 tRNA 如何在 A 位点定位，A 位点定位于 30 S 与 50 S 亚基之间。

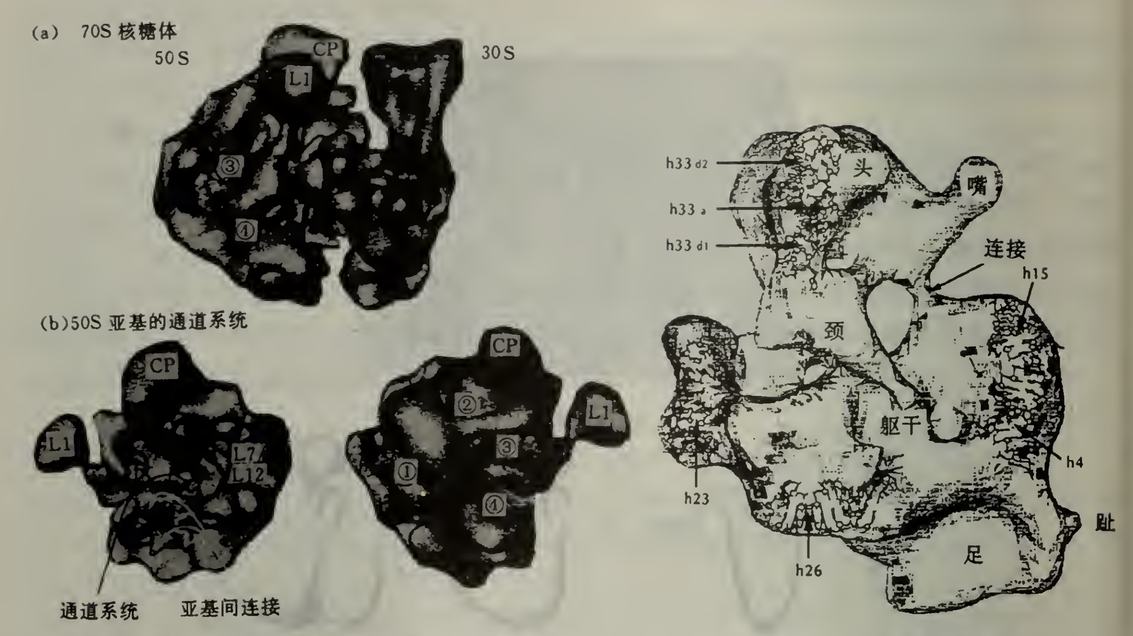


图 9-6　大肠杆菌 70 S 核糖体的三维重组

(a)从 50 S 亚基 L 1 突起一侧去看 70 S 核糖体；(b)除去 30 S 亚基大部分可看出亚基间的连接与 2 个通道系统的入口(黑箭头)。

图9-7　30 S 亚基半透明图，显示 16 S rRNA 的装配位置

RNA 各螺旋的编号用"h"表示。

图 9-8 示 2 个 tRNA 分子在核糖体上的装配。在核糖体的 30 S 和 50 S 亚基之间有"桥"作为主要的连接，在这个桥的上方有亚基间空腔，这个空腔刚好可以容纳在 A 位点和 P 位点的 2 个 L 形的 tRNA 分子。这 2 个 tRNA 分子呈"S"形构型排布，即 A 位点的 tRNA 的 D-环面向



图 9-8　2 个 tRNA 装配入核糖体，2 个分子呈"S"构型排布

(a)70 S 核糖体(半透明)tRNA 的反密码子环(anticodon loop)在左侧，而 CCA 末端则在右侧，P 位点的 tRNA 的一部分在界面的桥后面；(b)70 S 核糖体(不透明的)的俯视图，反密码子环(埋在 30 S 亚基内)是在左侧，而 CCA 末端(埋在 50 S 亚基内)则是在右侧。

P 位点 tRNA 的 T-环，这便使得它们可以合适地装配入桥的转折处。这样的排布使tRNA的反密码子环与在 30 S 亚基的头部与躯干之间的 mRNA 接触，而 tRNA 的 CCA 末端则在出口通道的入口处附近。

70 S 核糖体大约是一个半径为 11 nm 的球体，由此算出其体积应约为 5 600 nm^3。而根据核糖体的分子量计算，其体积则约为 2 600 m^3，占结构的 45%，其余的结构 55%是由溶剂所组成，那些与溶剂有关的部分则包括：亚基间的间隙，以及全部洞穴与通道。这样，核糖体实际上是一个具有凹陷、空洞、裂隙和坑道的结构。

二、核糖体核糖核酸（rRNA）

核糖体核糖核酸（rRNA）是核糖体的组成成分。它们对于核糖体的自身组装和活力表现起着重要作用。综合运用现代生物学、化学和物理学技术测定，得知每一个细菌核糖体中均有 2 个大的 rRNA（16 S，23 S）和一个小的（5 S）rRNA 分子。它们是整合的组分，除去了它们便会使核糖体结构完全瓦解。由此可知，rRNA 的功能是和核糖体蛋白质相互作用以维持核糖体的三维结构。另外，rRNA 还直接参加 mRNA 与核糖体小亚基的结合以及亚基间的联合。体外重组实验表明，核糖体蛋白质本身无蛋白质合成活性，缺少部分核糖体蛋白质也不会导致核糖体的失活，rRNA 在蛋白质合成中起着决定性的作用，这一点已被越来越多地研究所证实。

在核糖体中，小亚基中的 16 S rRNA 链长为 1 542 个核苷酸；作为大亚基的一个组分的 23 S 则含有 2 904 个核苷酸。此外，每一个大亚基还有一个很短的 5 S rRNA，含有 120 个核苷酸。所有 3 种 rRNA 均为单链，其鸟嘌呤和胸腺嘧啶以及腺嘌呤和尿嘧啶均不相等。虽然如此，还是有足够等价的碱基对，使得在同一条链上许多碱基对形成氢键，生成像在 tRNA 中的发夹结构。rRNA，特别是大的 rRNA 分子中，有多个螺旋区和环区组成的区域，每个区域大概是在结构上和功能上相对独立的单位。

（一）16 S rRNA

1987 年第一次将大肠杆菌的 16 S rRNA 的 1 542 个核苷酸序列确定之后，根据序列的互补性以及 RNA 序列的某些部分对核糖核酸酶和化学试剂的敏感性，提出了一个折叠的形式，并且得知不同来源的 16 S rRNA 都具有相似的二级结构。后来发现真菌细胞 18 S rRNA 和线粒体 12 S rRNA 也有类似的二级结构（secondary structure）模式。

大肠杆菌 16 S rRNA 的二级结构见图 9-9。1 542 个核苷酸中已知有 10 个核苷酸被甲基化，例如，m^7G_{966}，m^5C_{967}，$m^2C_{m\,1\,402}$，$m^5C_{1\,047}$，$m_2^6A_{1\,518}$，$m_2^6A_{1\,519}$等。其中 N^6，N^6-二甲基 A 及其相邻序列参与 mRNA 与 tRNA 的相互识别。由非甲基化的 16 S rRNA 构建的核糖体，在体外只能低效地进行蛋白质的合成；另一方面，缺少 $m_2^6A_{1\,518}$与 $m_2^6A_{1\,519}$修饰碱基时，核糖体对春日霉素表现出抗性。

16 S rRNA 在蛋白质生物合成的起始、延伸与终止起着活性作用，它的反 Shine-Dalgarno 序列通过与 mRNA 上的 Shine-Dalgarno 序列碱基配对而与 mRNA 结合。这 Shine-Dalgarno 序列位于 mRNA 的起始密码子 AUG 上游附近。这种相互作用对维持 mRNA 的正确阅读框架无疑是重要的。

解码位点，即密码子与反密码子发生相互作用的位点，位于 16 S rRNA 的 1 400 位置的 C 残基的 1 或 2 nm 之内。当 tRNA 位于核糖体 P 位点时，利用紫外光可以使 tRNA 反密码子的

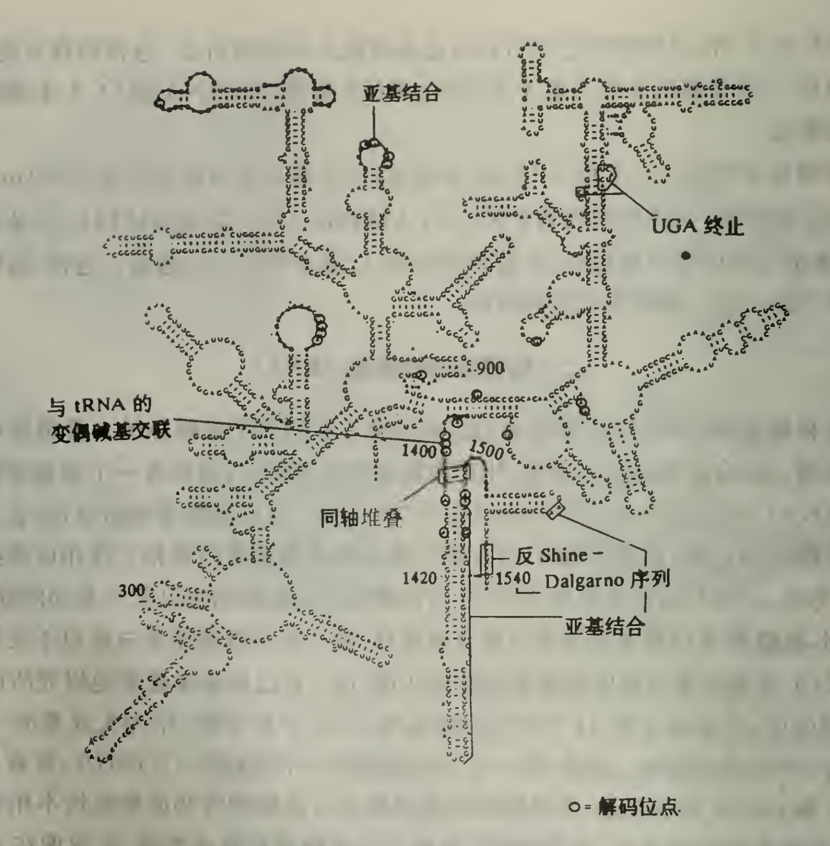


图 9-9　大肠杆菌 16 S rRNA 的功能区

5′(变偶)碱基与 C 残基交联。围绕在 1 400 与 1 500 位置的这个高度保守核苷酸序列有助于形成解码位点。现认为 16 S rRNA 的 1 400～1 500 序列是“解密码区”的重要组成成分。1 404～1 445/1 496～1 497 间的配对可与 mRNA，tRNA 共同形成碱基堆叠达到稳定配对，提高翻译的准确性。16 S rRNA 的 C_{1400}可通过形成 5，6-环丁二嘧啶与 tRNA 结合。一级结构中不配对区域之间可能包括非 Watson-Crick 的相互作用。在 16 S rRNA 分子的一些茎区中还出现单碱基突起，如 G_{31}，A_{746}，A_{1441}，至少其中 2 个(G_{31}，A_{746})被认为与核糖体蛋白的识别有关。现在认为 16 S rRNA 的功能区域可能是位于核糖体的表面。

现知 rRNA 与核糖体的大小亚基结合有关，16 S rRNA 中一些碱基为亚基结合所必需。16 S rRNA 中 790 环是与 50 S 亚基的结合部位。如果 G_{791}突变为 A，即会减弱 30 S 与 50 S 亚基的结合。另一个与亚基结合有关的部位是 16 S rRNA 的 1 409～1 491 序列，这个茎结构的缺失或 G_{1416}颠换为 U，则亚基之间的结合就不能实现；而 A_{1518}和 A_{1519}的非甲基也会使亚基结合减弱。

在 UGA 终止区，C_{1200}和 C_{1203}突变则会减弱终止密码子 UGA 的终止效率。

(二)23 S rRNA 与 5 S rRNA

23 S rRNA 是由 2 904 个核苷酸残基组成。它折叠成部分双螺旋结构，其特征在进化上是保守的。23 S rRNA 的二级结构的重要功能区域见图 9-10。

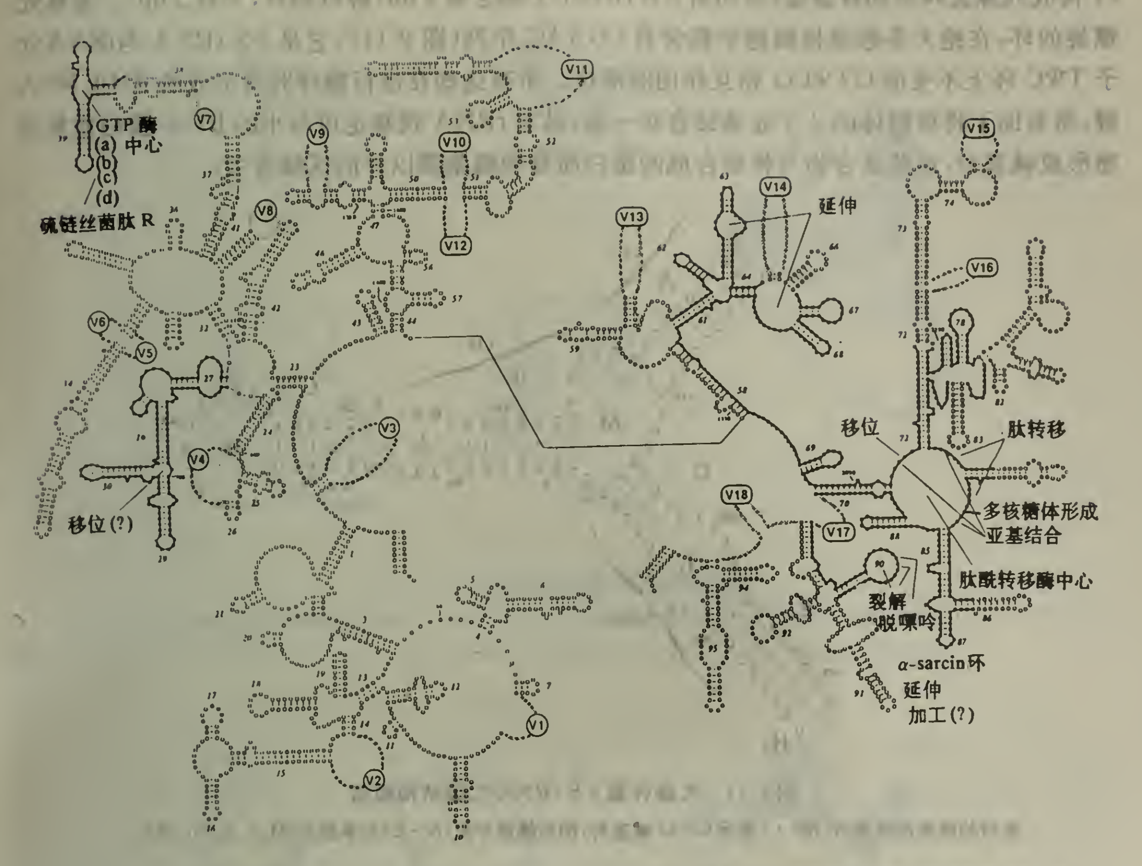


图 9-10　23S rRNA 功能区域的二级结构图解，一般地代表大亚基 rRNA

(a)起始；(b)延伸；(c)终止；(d)应激反应。

扩大(或不变)片段从 5′-末端起编号。

现发现所有 23 S rRNA 中，803～811 序列是高度保守的，并可能具有5′UAGCUGGUU3′序列。23 S rRNA 的三级折叠表明，肽酰转移酶的位点由 2 个区域组成，其中心围绕在 803～811 与 2 030～2 615 位置，而它的 $A_{2\,451}$，$C_{2\,452}$ 则是肽酰-tRNA 与 P 位点的结合位点。

延伸因子 EF-Tu 与 EF-G 彼此竞争在核糖体上的结合位点。EF-G 交联于 23 S rRNA 的 1 067位置区域。化学保护作用研究表明，EF-G 也与 2 660 位置附近的高度保守的核苷酸序列相互作用，这个位置是细胞毒素 α-sarcin 与蓖麻蛋白作用位点。

5 S rRNA 由 120 个核苷酸残基组成。5′-末端附近的序列与 3′-末端附近的序列互补，形成“Y”形二级结构(图 9-11)。由于并不进一步折叠，由此可知缺少三级的相互作用。

尽管 5 S rRNA 对核糖体活性是必需的，可是它的作用仍不十分清楚。很可能当核糖体执行功能时，它与 16 S rRNA 起相互作用，与已知的真核生物的 5 S rRNA 与 18 S rRNA 相互作用相类似。

试验指出，rRNA 中许多不配对的碱基与其它 RNA 分子结合于核糖体上有关。如靠近 16 S rRNA 3′端的多个不配对碱基，与 mRNA 分子起始位点 5′侧有一段不转录的前导序列互补，这无疑对蛋白质合成起始的正确与否极为重要；再有，在 5 S rRNA 中，由 31～34 和 48～

51 构成的螺旋叫共同臂螺旋，在所有 5 S rRNA 中都含有 4 bp（酵母例外，只有 3 bp）。紧靠此螺旋的环，在绝大多数原核细胞中都含有 CGAAC 序列（图 9-11），它是 5 S rRNA 与tRNA分子 TΨC 环上不变的 GTΨCG 相互作用的部位。所有这些在进行翻译时所形成的 RNA-RNA 键，将有助于将核糖体的 2 个亚基结合在一起，其它 rRNA 残基也可与小的 RNA-蛋白质复合物形成碱基对，这些复合物与将新合成的蛋白质导向细胞膜以外的区域有关。

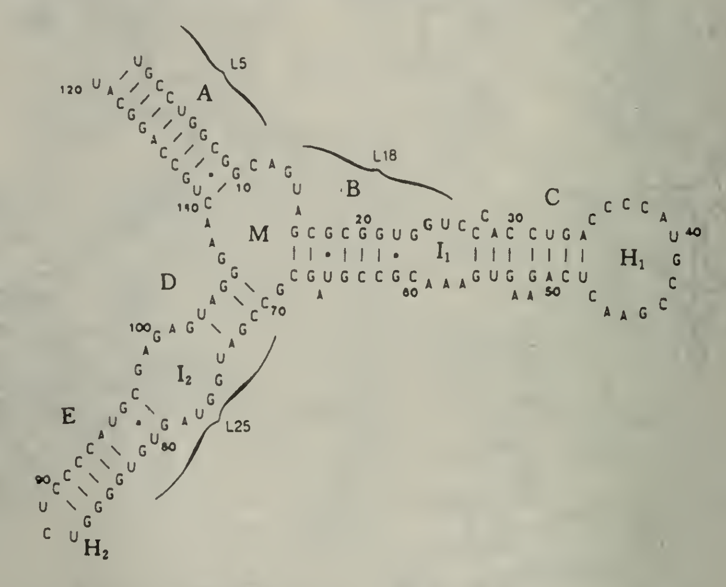


图 9-11 大肠杆菌 5 S rRNA 二级结构模型

配对的碱基用线表示；用“·”表示 G—U 碱基对；图示螺旋片段（A～E）与单链环（M，I_1，I_2，H_1，H_2）。

（三）核糖体小亚基模型

小亚基的模型见图 9-12。平台可能是密码子与反密码子相互作用的位点，通过区域包括有 16 S rRNA 的 3′-末端，N^6-二甲基腺嘌呤核苷酸 1 518 与 1 519，以及由 Shine-Dalgarno 序列为前导的 mRNA 起始密码子序列。蛋白质 S 6，S 11，S 15，S 18 也发现在平台上，它们与mRNA结合有关。蛋白质 S 3，S 10，S 14 与 S 19 则和依赖于聚（U）的 $tRNA^{Phe}$的结合有关。

（四）核糖体大亚基模型

大亚基具有一个明确的轮廓。具有 3 个特征：L 1 脊、中央突起、L 7/L 12 茎。中央突起是由蛋白质 L 27，L 18 与 5 S rRNA 组成。茎由 2 个 L 7/L 12 二聚体与位于它的底部的 L 10 组成（图 9-13（a）与（b））。肽酰转移酶活性位点定位于中央突起与脊之间。茎与移位有关。新生肽链的出口位点位于 P 位点对面，膜结合位点定位在蛋白质 L 19 的区域（图 9-13 c）。蛋白质在亚基内的排布见图 9-13（d）与（e）。

利用 X 射线衍射对 5 S rRNA 的研究证实，L 5，L 18，L 25 是与 5 S rRNA 的结合蛋白（图 9-11）。5 S rRNA 与 L 18 结合位点分别是 23，24，51，54，56，64，67 位的 G。L 18 的 18～25 氨基酸的序列（Leu-Gln-Glu-Leu-Gly-Ala-Thr-Arg-）是识别 5 S rRNA 的结构域。5 S rRNA 的 A_{66}，G_{67}与 L 18 的 Gln 19 间通过形成氢键使 5 S rRNA 与 L18 相互作用。核糖体组分的异源重组实验发现原核生物的 L 5，L 18，L 28 不能与酵母 5 S rRNA 结合，但它们的 L 18，L 25 能与酵母 5.8 S rRNA 结合。

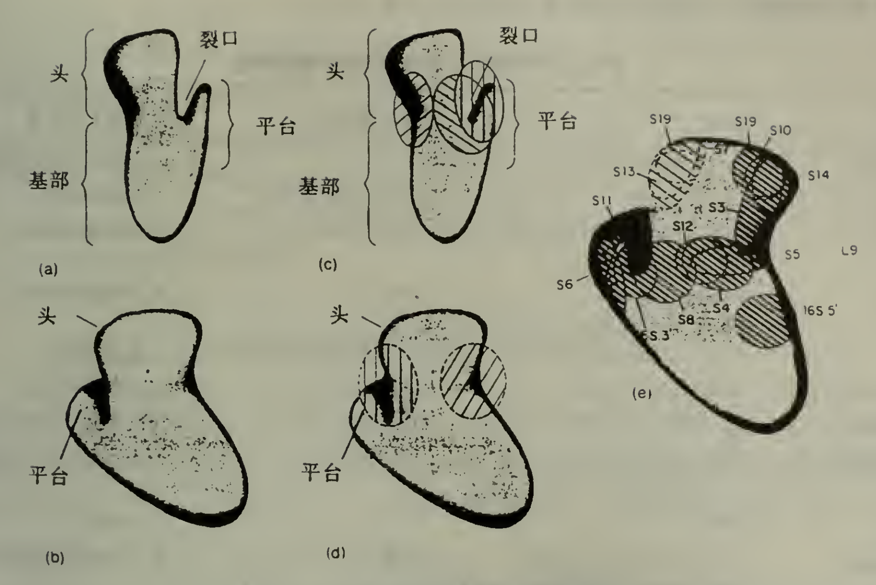


图 9-12 大肠杆菌核糖体小亚基模型

(a)与(b)是小亚基在常规电镜显示出的头、裂口、平台主要特征的模型;(c)与(d)显示出起始因子 IF1,IF2,IF3 以及延伸因子 EF-G 和 EF-Tu 的结合位点;(e)是 16 S rRNA 3′-末端与 5′-末端及一些小亚基蛋白质定位。

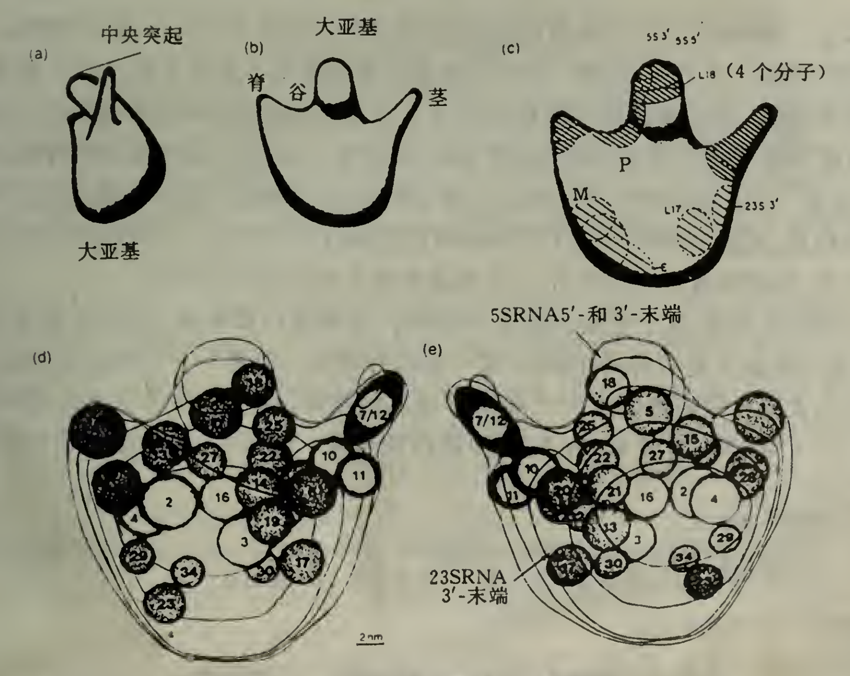


图 9-13 大肠杆菌核糖体大亚基模型

(a),(b):用常规电镜观察到的形状,显示出中央突起,L 1 脊和 L 7/L 12 茎;(c):几种大亚基蛋白、5 S rRNA,23 S rRNA 的功能区域的定位,P,M,E 分别代表肽酰转移酶位点、膜结合位点和新生肽链出口位点;(d),(e):表示大亚基蛋白质在大亚基内的排布。

现将有关核糖体主要结构功能域的特征列于表 9-2。

表 9-2 核糖体主要结构功能域的特征

功能域	rRNA	核糖体蛋白质	核糖体中的位置
mRNA 结合	16 S 3′端 SD 序列	S 21	30 S 亚基大叶与基底的接合部
R 位点	16 S/m^{7}G 526	S 3,S 10,S 14,S 19	小亚基外表面
A 位点	16 S/G 530,A 1 492,A 1 493,A 1 408,G 1 494	S 7,S 11,S 19,S 18	大小亚基交界处
P 位点	23 S/803～811,16 S/1 393～1 401	S 13,S 19	大小亚基交界处
解码区	16 S/C 1 400～A 1 500	S 1～5,S 9,S 12,S 18,S 21	小亚基裂隙处
起始因子结合位点	16 S/1 392～1 408	S 12,S 13,S 19～Ⅱ	近裂隙与平台处
GTP$_{ase}$中心	23 S/1 030～1 130	7/12,L 10,L 11	50 S 亚基 L 7/L 12 分子柄
肽酰基转移酶中心	23 S/Ⅱ,575～745 Ⅴ,2 030～2 615	L 2,L 16,L 23	中央突起与 L 1 突起之间
E 位点	23 S		50 S 亚基
亚基重组区	16 S/787～795,811～820 23 S/2 306～2 313,2 750～2 757	S 12 等	大小亚基的裂隙与凹槽

三、核糖体的自我组装

核糖体含 3 种 rRNA 和 53 种蛋白质，这些成分如何组装成具有活性的核糖体，是一个值得深入研究的问题。在 1968 年第一次完成大肠杆菌核糖体小亚基由其 rRNA 和蛋白质组分在体外的重新组装。这个重组装的颗粒具有与 30 S 亚基功能完全相同的蛋白质合成活性。这个重组装只需 16 S rRNA 和 21 种蛋白质，而不需要加入其它组分（如酶或特殊因子），表明是一个"自我组装"（self-assembly）的过程。所谓自我组装，是指进行组装所需要的全部信息来源都在亚基结构里，其蛋白质和 rRNA 都带有规定组装过程的信息。自我组装的驱动力包括疏水性相互作用、氢键和离子相互作用，以及碱基堆叠之间的相互作用等。

进一步的研究发现，这个组装过程有一定顺序，即某蛋白质的加入要先于其它蛋白质的加入。而且各组分的加入是有协同作用的，即一种组分的加入加强了下一种组分的加入。图 9-14 概略地表示出大肠杆菌核糖体的 30 S 和 50 S 亚基的组装过程。在 30 S 亚基的组装过程（图 9-1 4上）中，以16SrRNA为骨架，先与15种蛋白质结合（实际上，这些蛋白质的结合也有先

16 S rRNA + 21 S 蛋白质 —0℃→ 21 S 颗粒（含有 15 S 蛋白质）—40℃→ 26 S 颗粒（含有全部 21 S 蛋白质）—0℃→ 30 S 核糖体亚基

23 S rRNA + 5 S rRNA + 34 L 蛋白质 —0℃，4 mmol/L Mg^{2+}→ 33 S 颗粒（含有约 20 L 蛋白质）—44℃→ 41 S 颗粒（含有全部 34 L 蛋白质）—0℃，10 mmol/L Mg^{2+}→ 48 S 颗粒 —50℃→ 50 S 核糖体亚基

图 9-14 *E. coli* 核糖体 30 S，50 S 亚颗粒体外组装途径

图示需要输入能量（构象改变）的步骤，在 30 S 组装中，依赖 rRNA 制备方法，证实有 7～13 种 S 蛋白质可直接结合于 16 S rRNA；大约 17 种 L 蛋白质是直接结合于 23 S rRNA，3 种 L 蛋白质结合到 5 S rRNA。

后)，形成21 S颗粒。然后再加上其余6种蛋白质(相当于上文的脱落蛋白)，最后组装成30 S亚基。这个组装过程，可能在16 S rRNA转录开始之后即行发生，因为在rRNA链上结合最强的部位，都集中在最先转录的5′端。

大肠杆菌核糖体50 S亚基的组装比较复杂，它包括两种(23 S和5 S)rRNA，而且蛋白质数目也较多，图9-14下半部分概括地表示出这个组装过程。先由23 S rRNA，5 S rRNA与约20种L蛋白结合，生成33 S颗粒。然后再加上其余蛋白质，组装成41 S颗粒。最后在Mg^{2+}(0℃)和50℃下转变为50 S亚基。

第三节 氨基酸的激活与氨酰-tRNA的合成

氨酰-tRNA是氨基酸掺入到多肽链中的一个关键性的中间物质，并且它们的合成执行着双重机能。首先，tRNA的氨酰化作用需要ATP，它为肽键的合成提供能量。其次，它准许在核糖体上肽键形成时，mRNA密码子的3个碱基要和tRNA上与之互补的反密码子之间通过反向平行的碱基配对，执行遗传信息的解读过程。

现知原核生物与真核生物细胞质含有40多种不同的tRNA，线粒体与叶绿体含有的tRNA数量则少于此，并且结构上也与细胞质的有所不同。在动物与酵母线粒体中各自仅含有22与24种tRNA，可知，线粒体的DNA中，tRNA的基因也会相应地减少。

一、tRNA的结构

现在已从各种不同生物中测定350多种tRNA的核苷酸顺序。所有tRNA都是单链分子，长度大约是80个核苷酸残基。tRNA的二级结构呈三叶草形，由4个臂和4个环组成，其中可变环是含有5～21个核苷酸残基(图9-15 a，9-15 b)。用X射线结晶学阐明tRNA的三级结构是一个紧密的L-形分子，远离的残基间由氢键和疏水堆叠之间相互作用维持了稳定(图9-16)。

接受臂是由tRNA 5′-末端与3′-末端的碱基配对形成。所有tRNA 3′-末端的—CCA是不成碱基配对的，它在原核生物tRNA生物合成中有两种情况，一类其自身有CCA三核苷酸，它们位于成熟tRNA序列与3′端附加序列之间，当加工切除附加序列后便显露出来；另一类则是其自身并无CCA序列，它是在切除3′端附加序列后，由tRNA核苷酰转移酶(nucleotidyl transferase)催化进行，并由CTP与ATP供给胞苷酰基与腺苷酰基。对于真核生物，其CCA序列的形成，则与原核生物后者的形成相同。tRNA 3′-末端的A残基是氨酰-tRNA合成酶反应时接受氨基酸的部位，这样，在三级结构上氨基酸的接受位点就远离反密码子。这使得氨酰基靠近肽酰转移酶，而反密码子(anticodon)则与小亚基上的mRNA密码子配对。—CCA序列对原核生物延伸因子EF-Tu和GTP识别荷载的tRNA，以及随后将氨酰-tRNA运入核糖体的A位点起重要作用。—CCA序列的修饰，例如，插入C残基，或是用U代替倒数第二个C，都会减弱与EF-Tu的相互作用。由此想到，延伸因子对—CCA末端的空间结构，以及3′-末端的A与氨酰基团的—NH_3^+具有识别作用。

所有生物的tRNA所特有的结构特征是许多各异的修饰核苷，包括单纯的碱基或核糖残基的甲基化，以及非常复杂的取代(图9-15 b，表9-3)。所有修饰是发生在tRNA基因转录之后，一般都是碱基的改变，而不是碱基的置换。然而偶然会发生碱基的交换，例如，插入次黄苷

或是不常见的碱基 queuosine，取代了在多核苷酸前体中的腺嘌呤或鸟嘌呤。修饰核苷可能有下列功用：三级结构的稳定，有助于密码子-反密码子的相互作用，防止被不适当的氨基酸所错载，增加翻译效率和信实程度以及维持解读的框架（图 9-16）。对缺少 tRNA 修饰酶的突变种研究发现，在邻近反密码子 3′侧（位置 37）的修饰核苷酸的改变，或是反密码子本身变偶位置（位置 34）核苷的改变，都显示出特殊的重要性。王德宝的研究证实，$tRNA_r^{Ala}$ 中，以 A 或 G 代替 I_{34} 的类似物，能接受丙氨酸，但所携带的丙氨酸掺入蛋白质中去的程度分别是原活力的 1/3 或 90% 以上，而如果以 A，G 或 C 代替 $tRNAr^{Ala}$ m^1I_{37} 的类似物时，也能接受丙氨酸，但掺入蛋白质中去的活力却只有 20%～30%。其它部位的修饰核苷，例如，真细菌的 tRNA 中位置 38，39 和 40 的假尿苷的改变，则会影响延伸的速度以及翻译的错误程度。

线粒体与叶绿体含有对其蛋白质合成所需要的全部 tRNA。线粒体 tRNA 含有 59～75 个核苷酸残基，比细胞质 tRNA 较小一些。哺乳动物线粒体仅含有单个基因编码起始与延伸的线粒体 $tRNA^{Met}$，显然，最初的转录本是由不同的途径来修饰，得出两种线粒体 $tRNA^{Met}$。而一般它们是由不同的基因来转录的。

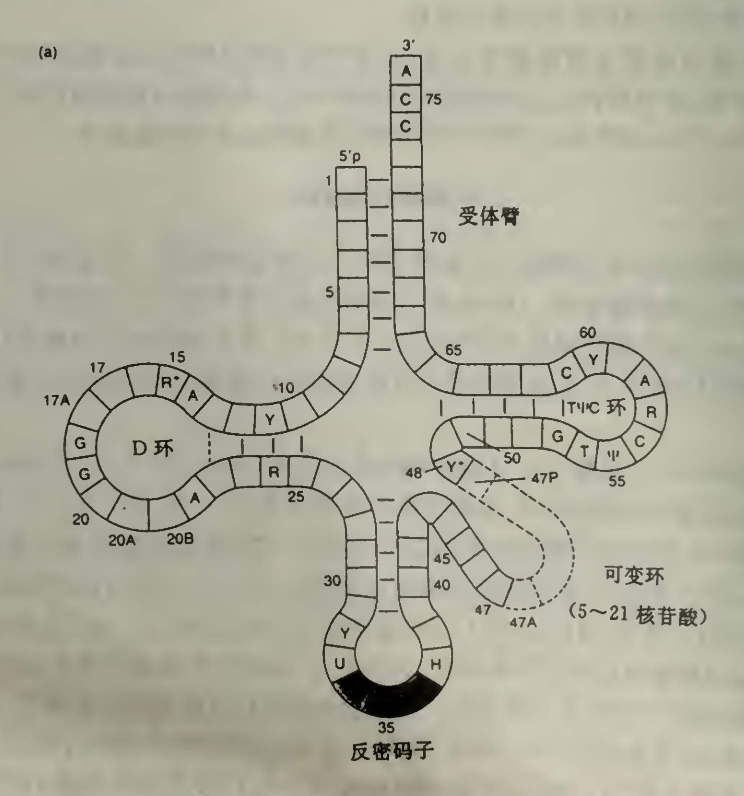


图 9-15 a　转移 RNA 的共同的二级结构

显示出除了起始 tRNAs 之外的所有 tRNA 序列中可变与半可变碱基的位置，用酵母 $tRNA^{Phe}$ 的编号，并对附加的碱基用下列字母表示：Y＝嘧啶，R＝嘌呤，H＝超修饰嘌呤，R^+ 和 Y^+ 常是互补的，位置 9 与 26 通常是嘌呤。位置 10 常是 G 或修饰的 G，可变环的大小是 5～21 核苷酸，D 环的大小上也有一定的可变性。

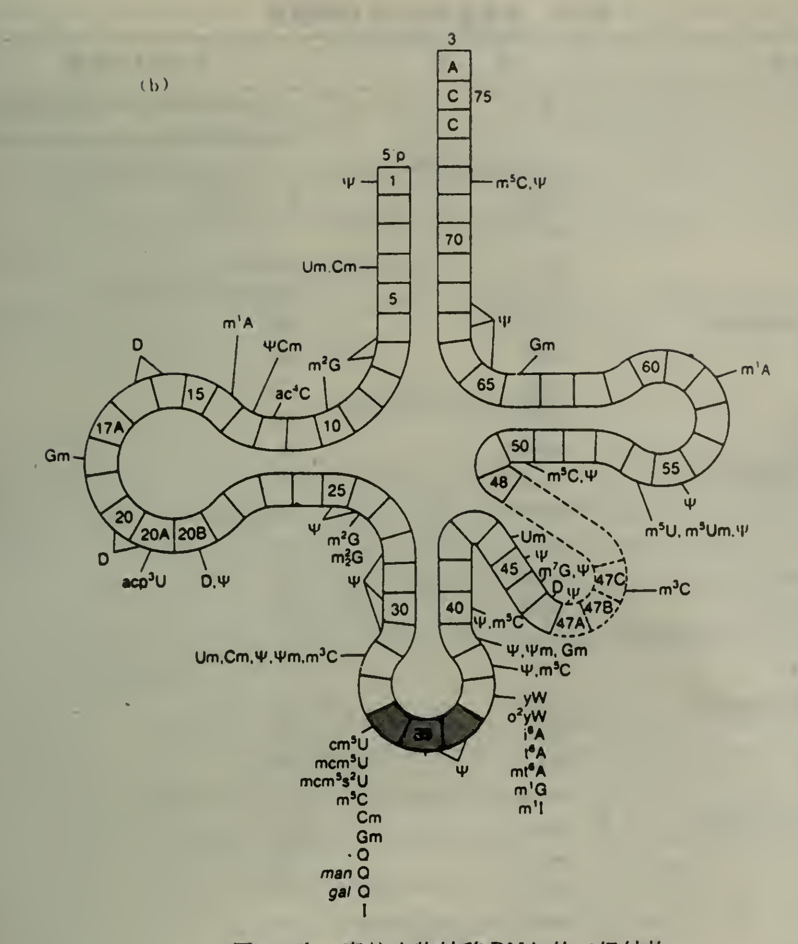


图 9-15b　真核生物转移 RNA 的二级结构

显示了在不同 tRNA 中修饰核苷的变化和其分布，最常见的修饰核苷是假尿苷(Ψ)，其它常见的修饰包括碱基或核糖 2′-位置的甲基化，缩写符号见表 9-3。

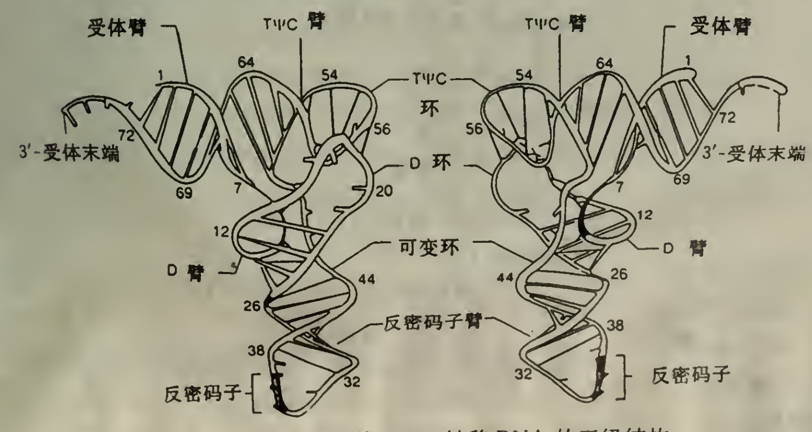


图 9-16　转移 RNA 的三级结构

酵母 $tRNA^{Phe}$ 的 2 个侧视图显示出核糖磷酸骨架，数字代表了序列中的核苷酸残基编号。碱基之间的氢键用横档表示，不成氢键的碱基用短杆附着在多核苷酸骨架上。

表 9-3 真核生物 tRNA 修饰核苷

修饰核苷名称	缩写	在 tRNA 的位置
假尿苷	Ψ	1,13,20B,25,26,28,30,31,32,35,36,38 40,45,46,47A,50,54,55,65,67,68,72
2′-*o*-甲基核苷	Um,Cm,Gm,Ψm	4,13,18,32,34,39,44,64
N^2-甲基鸟苷	m^2G	6,7,10,26
N^4-乙酰胞苷	ac^4C	12
N^1-甲基腺苷	m^1A	14,58
5,6-二氢尿苷	D	16,17,20,20A,20B,47
3-(3-氨基-3-羧基・丙基)尿苷	acp^3U	20A
N^2,N^2-二甲基鸟苷	m^2_2G	26
3-甲基胞苷	m^3C	32,47C
5-氨甲酰基甲基尿苷	cm^5U	34
5-甲氧基羰基甲基尿苷	mcm^5U	34
5-甲氢基羰基甲基-2-硫尿苷	mcm^5s^2U	34
5-甲基胞苷	m^5C	34,40,50,72
queuosine	Q	34
β-D-manosylqueuosine	manQ	34
β-D-galactosylqueuosine	galQ	34
次黄苷	I	34
Wybutosine(以前为 Y)	yW	37
Wybutoxosine(过氧 wybotosine)	o^2yW	37
N^6-异戊烯腺苷	i^6A	37
N^6-苏氨酸羰酸腺苷	t^6A	37
N^6-甲基 t^6A	mt^6A	37
N^1-甲基鸟苷	m^1G	37
N^1-甲基次黄苷	m^1I	37
2′-*o*-甲基假尿苷	Ψm	38
N^7-甲基鸟苷	m^7G	46
5-甲基尿苷(核胸苷)	m^5U	54
2′-*o*-甲基-5-甲基尿苷	m^5Um	54

上表列出发现在不同真核生物的 tRNA 的修饰核苷的名称与缩写,以及它们出现的位置(见图 9-15 b)。

Q 衍生物

Q 或 Quo:R=H

(queuosine)

manQ:R=*β-D* 甘露糖基

galQ:R=*β-D* 半乳糖基

Wye 衍生物(前作 Y 衍生物)

Wyo=Wyosine(前作 'Yt')R=H

Wybutosine(前作 'Y')

R=—CH_2—CH_2—CH(—$COOCH_3$)—NH—$COOCH_3$

Peroxywybutosine(前作 'Yr. Yw 或 过氧 Y')

R=—CH_2—CH(—OOH)—CH(—$COOCH_3$)—NH—$COOCH_3$

二、氨基酸的激活

现已知氨基酸是不能直接与模板相结合的，氨基酸在被转运到模板之前必须与接合体相连接，这个接合体就是上文所述的 tRNA。氨基酸与 tRNA 连接形成氨酰-tRNA 即是氨基酸的激活，它是在细胞质内进行的。每一种氨基酸以共价键连接于一种专一的 tRNA，这个过程需要消耗 ATP，形成的氨酰键是一个高能键，使生成的复合物被激活。这个高能氨酰-tRNA 的键能（氨酰键）可用于较低能量的肽键的形成。

将氨基酸接合于 tRNA 以形成氨酰-tRNA 的激活反应是在氨酰-tRNA 合成酶的催化作用下进行的，下文将对此酶的性质加以阐述。

三、氨酰-tRNA 合成酶

氨酰-tRNA 合成酶的作用是将氨基酸接合于 tRNA，因此，它必须同时能够专一地与氨基酸的侧链基团以及与 tRNA 相结合。原核生物含有 20 种氨酰-tRNA 合成酶，每一种合成酶对一种氨基酸专一，但可以和该氨基酸的多个同工受体 tRNA 结合。但也有例外，如在大肠杆菌中只有一种 $tRNA^{Lys}$，但却有 2 种赖氨酸的氨酰-tRNA 合成酶。在真核生物内，细胞质、叶绿体和线粒体内的氨酰-tRNA 合成酶是不同的。现已对多种不同来源的合成酶进行提纯，并对 22 种酶进行了氨基酸测序。

各种氨酰-tRNA 合成酶的四级结构有很大差异，可以是单体（α）、二聚体（dimer）（α_2）和同型或异型四聚体（α_4 或 $\alpha_2\beta_2$），其多肽链长度为 300～900 个氨基酸残基，其中一些较长的肽链是由较短的肽链在氨基端延长而成的，其延长部分似乎与其催化功能无关。高等真核生物的氨酰-tRNA 合成酶的一个特点是形成特殊的聚集物（aggregate），可以形成分子量高达 1 000 kD 的由 11 条多肽链组成的复合物；有些复合物结合于内质网，有些则游离于细胞溶质中。

表 9-4　两类氨酰-tRNA 合成酶

第一类	第二类
Arg	Ala
Cys	Asn
Gln	Gly
Glu	Asp
Ile	His
Leu	Lys
Met	Phe
Trp	Pro
Tyr	Ser
Val	Thr

由大肠杆菌分离出的氨酰-tRNA 合成酶按其初级和三级结构以及反应机制的差异可分为二类，每类有 10 种酶（表 9-4），这个分类也适用于其它经研究过的生物。它们都含有 3 个区域，即催化域（ATP 和氨基酸结合位点）、tRNA 受体螺旋结合域和 tRNA 反密码子结合域。此外，多聚态的合成酶还有一个寡聚形成区域。催化域是一个大的区域，在其中插入 tRNA 受体螺旋结合域。第一类合成酶具有一个 N-末端催化域，它是一个称为核苷酸结合折叠（nucleotide-binding fold）的基序（motif），由平行的 β-折叠和 α-螺旋交替排列组成。在催化域中有两个短的氨基酸序列，称为“署名序列”（signature sequnce），形成 ATP-结合位点的一部分。tRNA 受体螺旋结合域以及酶的 C-末端的其它两个区域均视不同的酶而有很大差异。第二类合成酶的催化域有 3 个相同的序列，其活性位点含一个由多个 α-螺旋包围着的大的反平行 β-折叠。插入催化域的受体螺旋结合域的结构也视不同的酶而异。反密码子结合域在 N-末端。

氨酰-tRNA 合成酶催化的反应分两步进行。氨基酸先被氨酰-tRNA 合成酶活化，生成氨酰腺苷酸(AA-AMP)，其中氨基酸的羧基是以高能键连接于腺苷酸上，同时放出焦磷酸：

$$AA+ATP \xrightleftharpoons{\text{氨酰-tRNA 合成酶}} AA\text{-}AMP+PPi \tag{1}$$

在正常情况下，AA-AMP 中间物仍然紧密地结合在酶上。直到与该氨基酸专一的 tRNA 分子碰撞时为止。随后同一酶将氨基酸转移至 tRNA 的末端腺苷酸残基上。

$$AA\text{-}AMP+tRNA \xrightleftharpoons{\text{氨酰-tRNA 合成酶}} AA\text{-}tRNA+AMP \tag{2}$$

反应(1)与反应(2)加成后的总反应为：

$$AA+tRNA+ATP \xrightleftharpoons{\text{氨酰-tRNA 合成酶}} AA\text{-}tRNA+AMP+PPi \tag{3}$$

总反应(3)的平衡常数接近于 1，自由能降低极少。这说明 tRNA 与氨基酸之间的键是高能酯键，高能键的能量来自 ATP 的水解。这个键水解时的标准自由能变化为－30.51 kJ。由于反应中形成的 PPi 水解成正磷酸，对每个氨基酸的活化来说，净消耗的是 2 个高能磷酸键。因此，此反应是不可逆的。

上述二步反应的机制如图 9-17 所示。可以看出，反应的第二步因两类酶而不同，第一类酶先将氨酰基转移至 tRNA3′端腺苷酸的 2′-OH 上，然后通过转酯作用转移至 3′-OH 上；第二类酶则直接将氨酰基转移至 3′-OH 上。

蛋白质合成的一个关键问题是如何保证 20 种不同的氨基酸按 mRNA 上遗传密码的顺序正确地掺入到多肽链上去，以形成正常的具有生理活性的蛋白质。为此，氨酰-tRNA 合成酶必须能正确识别其底物氨基酸和 tRNA，如发生错误，则必须具有进行校正的机制。

不同氨酰-tRNA 合成酶对氨基酸的专一性是不同的，有些是高度专一的，只与一种氨基酸结合，有些则同时能与正确氨基酸结构相近的氨基酸结合。但虽然有时能结合上这些类似的却是不正确的氨基酸，最终却不能生成稳定的氨酰-tRNA，这是因为存在进行校正(proofreading)的机制。校正包括两个阶段，这两个阶段均需要相关 tRNA(cognate tRNA)参与。第一阶段的校正是当相关 tRNA 结合上时，即将已生成的不正确的氨酰-AMP 水解掉，如甲硫氨酸、异亮氨酸、缬氨酸的氨酰-tRNA 合成酶即是如此。第二阶段的校正是当不正确的氨基酸转移至 tRNA 上时，由于 tRNA 在其结合位点识别其结构错误而将之水解掉。这两个校正阶段如图 9-18 所示。上述两个阶段的校正均要求相关 tRNA 的参与。甚至在形成氨酰-AMP 之前，tRNA 也起着引发校正的作用。例如，大肠杆菌的 Ile-tRNA 合成酶同时可以催化缬氨酸与 AMP 结合，但当加入 $tRNA^{Ile}$时，Val-AMP 便发生水解。这说明相关 tRNA 在校正氨基酸结合中的重要性。

氨酰-tRNA 合成酶也必须识别正确的 tRNA。现已知合成酶蛋白是在 L 形 tRNA 的侧面与之结合的，而且两类合成酶结合的侧面不同，它们是分别在相对的侧面结合的(图 9-19)。具体地说，第一类酶(如Gln-tRNA合成酶)是在tRNA的D-环侧结合，识别其受体臂的小沟(minor groore)，反密码子环在另一端。而第二类酶(如 Asp-tRNA 合成酶)则在另一侧与tRNA接触，识别其可变环和受体臂的大沟(major groove)。由于 tRNA 的受体臂和反密码子臂(anticodon arm)是和合成酶紧密接触的，因此，对 tRNA 的识别也是在这两部位。利用突变的方法，改变tRNA的个别碱基，然后测定合成酶对它的识别能力，可以了解哪些核苷酸是与

氨基酸 + ATP

腺嘌呤

PP_i

5′-氨酰腺苷酸
（氨酰-AMP）

tRNA的3′端

腺苷

氨酰-AMP

tRNA

AMP

第一类酶

转酯作用

第二类酶

氨酰-tRNA

图 9-17　氨酰-tRNA 合成酶的 tRNA 氨酰化作用

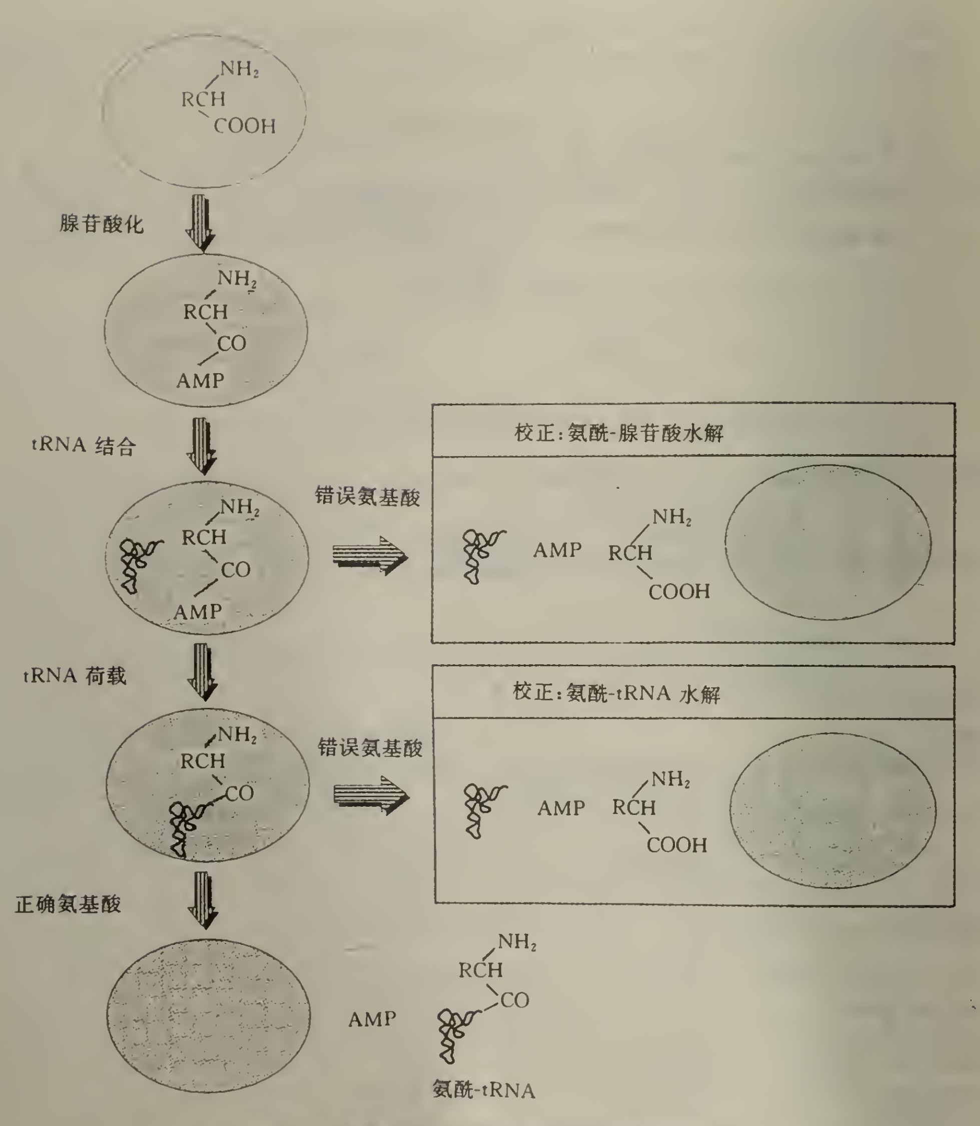


图 9-18　氨酰-tRNA 合成酶结合错误氨基酸的校正

识别有关的。表 9-5 示已知的一些 tRNA 的识别位置。据现在所知，这些识别位置为数不多，一般只有 1～5 个。如 Ala-tRNA 合成酶的识别位置只是在受体臂上的 G3 · U70 bp 上。而且，每一种合成酶对 tRNA 的识别位置均不相同，没有一般的规律。

由于氨酰-tRNA 合成酶对 tRNA 的识别在蛋白质的正确合成中的重要性，因此，有人称之为“第二遗传密码”。但是，这“第二遗传密码”(secondary code)的具体内容如何，现仍未清楚。显然，它是比经典的遗传密码(见第八章)复杂得多的。



第一类



第二类

图 9-19　第一类和第二类氨酰-tRNA 合成酶的结晶结构

示 tRNA（深色）分别结合于酶蛋白（浅色）的相对侧面。

表 9-5　一些 tRNA 中对识别合成酶起决定作用的核苷酸

tRNA	起作用的核苷酸
大肠杆菌丙氨酸	G3・U70
大肠杆菌精氨酸	A20 和其它
大肠杆菌谷氨酸	U35
大肠杆菌异亮氨酸	L34
大肠杆菌甲硫氨酸	反密码子
大肠杆菌丝氨酸	G1・C72、G2・C71、A3・U70、C11・G24
大肠杆菌缬氨酸	反密码子
酵母苯丙氨酸	G20、G34、A35、A36、A73

第四节　原核生物的蛋白质的生物合成

一、核糖体的作用

在细胞质中，大多数 mRNA 是与一个或数个核糖体结合于称为多核糖体的复合物中（图 9-3）。核糖体在多（聚）核糖体上的数目受 mRNA 密码子序列的长短、以及在核糖体循环中起始、延伸、终止的相对速度所控制。例如，合成珠蛋白链的 mRNA 密码子序列含有约 450 个核苷酸，其多核糖体最大是五聚体。这表明两个连续的核糖体中心之间的平均距离是 80～90 个核苷酸。如果蛋白质合成的起始相对慢于延伸和终止，核糖体沿着 mRNA 将会有更宽阔的空间，结果是核糖体少。相反，起始比延伸或终止快，mRNA 上组装的核糖体紧密，导致多量的核

糖体。一般来说，特定的 mRNA 的翻译效率是受起始速度所决定。在蛋白质合成的调控中，起始作用的控制是一个重要因素。当起始作用被抑制，例如，被金红三羧酸(aurintricarboxylic acid)或密旋霉素(pactamycin)抑制之后，已经启动蛋白质合成的核糖体是仍能完成肽链的延伸、终止，并从 mRNA 上解离。

核糖体由容易解离的亚基组成，表明其在蛋白质合成的某个阶段大亚基与小亚基有周期性的分离。将细胞放在重同位素中生长，然后转入轻介质中，不久即开始出现杂交的核糖体(重 50 S/轻 30 S 和轻 50 S/重 30 S)，它们出现的速率表明，多肽合成的每周转一次就有一次亚基交换。体外蛋白质合成研究也证实这个推断：在有大量过剩的轻核糖体存在时，约 1 min 左右(这是合成一条多肽链所需的时间)，差不多全部重核糖体都消失，同时出现杂交的核糖体(图 9-20)，用抗菌素稀疏霉素(sparsomycin)做体外试验证实了亚基交换对蛋白质合成的专一性依赖关系。这种化合物抑制了多肽链的延长，也防止了亚基的交换。

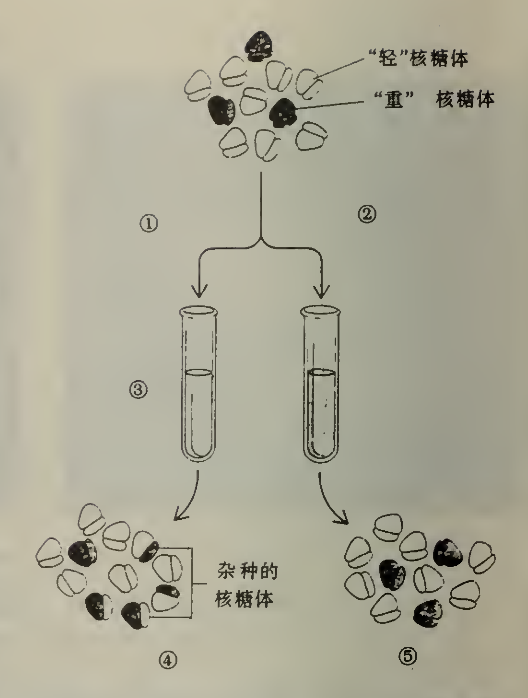


图 9-20 亚基交换对蛋白质合成专一性依赖

①将一些"重"核糖体在存有过量"轻"核糖体时加入含有蛋白质合成所需全部组分的试管中；②做类似的实验但此时加入稀疏霉素。这个化学试剂阻碍蛋白质的合成；③短时间温育后，将核糖体分离；④蛋白质合成时，分离出不少的杂种核糖体(即重 30 S/轻 50 S 或轻 30 S/重 50 S)：发生了亚基交换；⑤蛋白质合成受阻时，没有亚基交换，分离不出杂种的核糖体。

二、mRNA 的结构与功能

所有的 mRNA 都是单链的多聚核苷酸，含有一个决定氨基酸序列的编码区，还有与起始密码子和终止密码子紧邻的侧翼序列。真核生物与原核生物的 mRNA 的区别在于前者在 5′端有一个帽子结构和在 3′端有一个多聚(A)的尾部，帽子结构由 7-甲基鸟苷和三磷酸与 mRNA的 5′端连接(图 9-21)。

在 mRNA 的编码区，二级结构对翻译速度影响较小。在多顺反子 mRNA 中，二级结构能调节个别解读顺反子的频率。二级结构的 5′-和 3′-不翻译序列对特定蛋白质的合成，例如，组蛋白、铁蛋白、微管蛋白和生长因子等的翻译调控可能会有特殊的重要性。

原核生物多顺反子 mRNA 常常有一个 5′-前导序列，以 pppA 或 pppG 开始，还有不翻译的顺反子间序列，以及在最后的顺反子终止密码之后有一个 3′-不翻译区域。前导序列的长度依 mRNA 而定，一般都是短的(约 20～200 个核苷酸)。也有些例外，其起始密码子位于 mRNA 5′-末端或距 5′端仅有几个核苷酸。原核生物转录终止之前，新生的 mRNA 即开始翻译，并且在前导序列上的核糖体能够改变新生 mRNA 的二级结构，从而调节操纵子剩余部分的转录。在某些情况下，前导序列也可以翻译成短的无功能的多肽。

顺反子间序列通常是短的，但也可以长达 400 个核苷酸。也有重叠顺反子的例子，它的

UGA 终止密码子的最后一个核苷酸就作为邻近起始密码子 AUG 的第一个碱基。

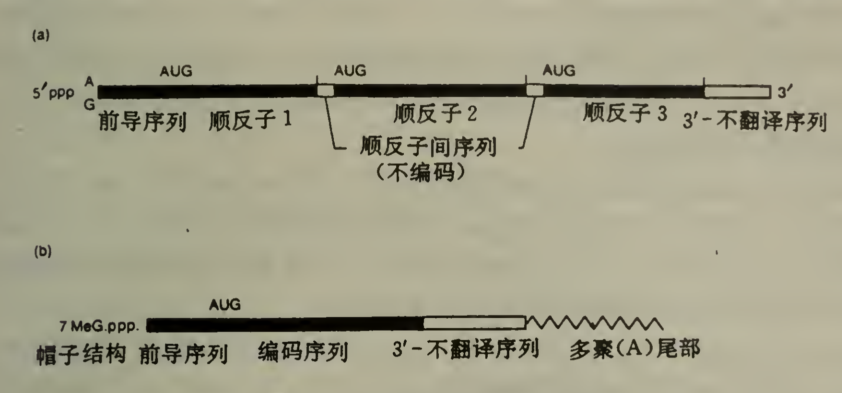


图 9-21　典型的信使 RNA 结构

(a)原核生物多顺反子 mRNA；(b)真核生物 mRNA。

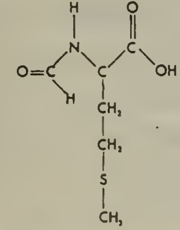
在蛋白质合成中，mRNA 3′-不翻译区域的功能还不清楚。它的主要作用可能是终止 RNA 聚合酶的转录作用，但也可能起 mRNA 稳定与翻译调控(translation regulation)的作用。

三、起始复合物的形成

(一)N-甲酰甲硫氨酸

在核糖体上进行的蛋白质的合成是从氨基末端开始，逐步加上一个个氨基酸，在羧基端终止。这可以用放射性标记的氨基酸加入合成血红蛋白的网织红细胞，然后立即分离新合成血红蛋白链得以证明。如果标记时间比完成一条链所需要的时间短的话，放射性很少会出现在氨基末端的氨基酸中，而大多是出现在羧基末端的氨基酸中，并且沿着从氨基端至羧基端可以观察到一个清楚的放射性增加的梯度。

所有细菌蛋白质合成的氨基端的第一个氨基酸都是 N-甲酰甲硫氨酸(fMet)，这是一个修饰了的甲硫氨酸，在其氨基末端连接上一个甲酰基(图 9-22)。一个像这样封闭的氨基酸只能用于蛋白质合成的起始阶段。由于不存在游离的氨基，这便防止它在肽链延伸时插入内部。这个甲酰基是在甲硫氨酸连接于 $tRNAf^{Met}$接合体上以后，由酶促加上去的。甲酰基的供体是 N^{10}-甲酰四氢叶酸，催化该反应的酶叫做转甲酰酶。



N-甲酰甲硫氨酸(fMet)

甲硫氨酸 (Met)

图 9-22　N-甲酰甲硫氨酸与甲硫氨酸结构的比较

$$\text{Met-tRNA}_f^{Met} \xrightarrow[N^{10}\text{-甲酰四氢叶酸}]{\text{转甲酰酶}} \text{fMet-tRNA}_f^{Met}$$

并非所有甲硫氨酰-tRNA 分子都可以甲酰化。大肠杆菌细胞内有两种类型 $tRNA^{Met}$，一类是 $tRNA_f^{Met}$，一类是 $tRNA_m^{Met}$。只有 $tRNA_f^{Met}$ 上的甲硫氨酸可以甲酰化，对 $tRNA_f^{Met}$ 和 $tRNA_m^{Met}$(内部甲硫氨酸)做顺序分析显示出，两者具有相同的反密码子顺序，但它们编码的氨基酸不同。大肠杆菌的 $tRNA_f^{Met}$ 和 $tRNA_m^{Met}$ 的结构主要有 3 点不同：①氨基酸臂 3′端第 5 个碱基在 $tRNA_f^{Met}$ 是 A，它与 $tRNA_f^{Met}$ 5′端 C 不配对；在 $tRNA_m^{Met}$ 相对应的位置是 C，它可以

与5′端的G形成配对；②TΨC环上，$tRNA_f^{Met}$是TΨCA，在$tRNA_m^{Met}$相对应位置是G；③反密码环上，$tRNA_f^{Met}$反密码子3′端邻位碱基是A，在$tRNA_m^{Met}$相对应位置是烷基化的A。实验发现，只有甲酰甲硫氨酰-$tRNA_f$能够和蛋白质的起始因子及30 S核糖体亚基结合，形成起始复合物，而甲硫氨酰-$tRNA_m$只能与延伸因子结合，将甲硫氨酸掺入肽链的中间。

(二)起始密码子的正确选读

细菌内蛋白质合成的起始是从核糖体小亚基30 S与fMet-$tRNA_f^{Met}$和一个mRNA分子形成复合物开始的。然后50 S亚基参加进去，形成有功能的70 S核糖体。每一个mRNA上有一个与核糖体结合位点，以合成一条独立的多肽链。每一位点内的核苷酸的顺序是使mRNA分子在蛋白质合成开始之前在核糖体上先行正确定位。

核糖体是如何辨别在一个基因开始的AUG密码子和编码内部甲硫氨酸的密码子呢？实验发现：将30个核苷酸长度的大肠杆菌噬菌体R17合成A蛋白的起始区mRNA片段，使其与大肠杆菌多核糖体形成起始复合物。然后用大肠杆菌素E3(colicin E3)(它可以从16 S RNA的3′-末端切下59个核苷酸)加到复合物中。用表面活性剂处理混合物，解离核糖体释放出RNA。释放出的RNA是mRNA-rRNA杂交分子。在这杂交分子内，mRNA和rRNA片段通过氢键结合在一起。

早在1975年，Shine等就注意到上述现象，由此提出一种很吸引人的假设来解释起始密码子的识别。他们见到几种细菌16 S rRNA 3′-末端顺序为：5′-PyACCUCCUA-3′。其中Py可以是任何嘧啶核苷酸。它可以和mRNA中离AUG顺序5′侧约10个碱基处有一段富含嘌呤的间隔顺序AGGA或GAGG(后来称此区域为Shine-Dalgarno顺序)互补。现认为正是由这样的配对将AUG(或GUG，UUG)密码子带到核糖体的起始位置上(图9-23)。

Qβ 噬菌体蛋白A	CUG	AGU	AUA	AGA	GGA	CAU	AUG	CCU	AAA	UUA
Qβ 噬菌体外壳	CUU	UGG	GUC	AAU	UUG	AUC	AUG	GCA	AAA	UUA
Qβ 噬菌体复制酶	UUA	CUA	AGG	AUG	AAA	UGC	AUG	UCU	AAG	ACA
λ 噬菌体Cro	AUG	UAC	UAA	GGA	GGU	UGU	AUG	GAA	CAA	CGC
f1 噬菌体外壳	UUU	AAU	GGA	AAC	UUC	CUC	AUG	AAA	AAG	UCU
ΦX 噬菌体174A	AAU	CUU	GGA	GGC	UUU	UUU	AUG	GUU	CGU	UCU
ΦX 噬菌体174A˙	UUG	CUG	GAG	GCC	UCC	ACU	AUG	AAA	UCG	CGU
ΦX 噬菌体174B	AGG	UCU	AGG	AGC	UAA	AGA	AUG	GAA	CAA	CUC
ΦX 噬菌体174E	GCG	UCG	AGG	CUU	GCG	UUU	AUG	GUA	CGC	UGG
脂蛋白	AUC	UAG	AGG	GUA	UUA	AUA	AUG	AAA	CGU	ACU
RecA	GGC	AUG	ACA	GGA	GUA	AAA	AUG	GCU	AUC	G
GalE	AGC	CUA	AUG	GAG	CGA	AUU	AUG	AGA	GUU	CUG
GalT	CCC	GAU	UAA	GGA	ACG	ACC	AUG	ACG	CAA	UUU
LacI	CAA	UUC	AGG	GUG	GUG	AAU	GUG	AAA	CCA	GUA
LacZ	UUC	ACA	CAG	GAA	ACA	GCU	AUG	ACC	AUG	AUU
核糖体L10	CAU	CAA	GGA	GCA	AAG	CUA	AUG	GCU	UUA	AAU
核糖体L7/L12	UAU	UCA	GGA	ACA	AUU	UAA	AUG	UCU	AUC	ACU
RNA聚合酶β亚基	AGC	GAG	CUG	AGG	AAC	CCU	AUG	GUU	UAC	UCC

16 S tRNA 3′-末端$_{OH}$ AUUCCU CCACUAG 5′

图9-23 已知部分蛋白质的mRNA在大肠杆菌核糖体上的结合位点

(蛋白质合成起始区域)

mRNA上互补于16 S tRNA 3′-末端的核苷酸用横线表示，应注意能与16 S tRNA 3′-末端碱基配对的长度与位置是可变的，与16 S tRNA 3′-末端相互作用的正确配对也是可变的。

进一步支持Shine-Dalgarno假设的证据来自用其它细菌核糖体进行R17A蛋白合成效率的研究。测定了6种细菌16 S rRNA分子3′末端顺序，发现碱基配对区的长度和强度(G·C对和A·U对的比值)是不同的。不同核糖体合成A蛋白的量与碱基对区的稳定性有关。表明起始频率的主要因素取决于mRNA-16 S rRNA碱基配对的强度。

虽然Shine-Dalgarno顺序是识别AUG(或GUG,UUG)起始密码子的基本条件，但并非充分条件。现知Shine-Dalgarno顺序会靠近不做密码子使用的AUG三联体。显然，为起始作用还必须有另一些mRNA的碱基顺序在起作用。在前导序列的上游序列，在mRNA与核糖体之间可能提供了识别信号，以保证这个Shine-Dalgarno序列在合适的构象内，或在整个核糖体的结合域内。整个mRNA分子的折叠以及30 S核糖体蛋白质(如S 1,S 4,S 18,S 21)对核糖体与mRNA的结合也起到某些作用。

多肽链的延伸一旦开始，16 S rRNA-mRNA的碱基配对便要在一定程度上解离，使松开的mRNA可以在核糖体表面上自由移动。当核糖体沿着mRNA移动时，它必须暂时地破坏双螺旋发夹区域(这些区域是许多游离mRNA区段的特征)，这样便创造出可以正确选择氨酰-tRNA前体的反密码子的单链区域。

(三)mRNA翻译的方向

当mRNA分子的起始区域正确地结合上一个核糖体后，在蛋白质合成中，它总是以固定的方向移动，即由其5′端向3′端的方向。开始解读的一端，即5′端，是最先合成的一端。这样，mRNA在DNA模板上合成过程中，核糖体可以连接在一条未完成的mRNA上。如果与此相反，多肽合成是朝3′→5′方向进行的话，那么，相应于一条完整的多肽链的mRNA则必须先完成，然后核糖体才能连接上去。蛋白质合成朝5′→3′方向，意味着在迅速生长的细菌细胞中，不与核糖体连接的长段mRNA通常是不存在的。

(四)起始因子

蛋白质合成的启动必须有起始因子的参加。起始因子是一类参与蛋白质生物合成起始的可溶性蛋白因子。蛋白质合成的起始要生成核糖体·mRNA·tRNA三元复合物，也叫起始复合物。复合物必须在起始因子帮助下才能完成。目前已知原核生物起始因子有3种。即:IF1，IF2和IF3。首先，这3种起始因子连接于30 S亚基上，GTP使这个结合稳定。GTP可能直接被IF2结合。IF1是一个小的碱性蛋白，它能增加其它2个起始因子的活性。有些原核生物不具相当于IF1的起始因子，在这些生物中，蛋白质合成的起始可以只在IF2和IF3参与下进行。现知IF1与16 S rRNA的结合位点分别是G_{529},G_{530},A_{1429}与A_{1493}。足迹分析表明IF1与氨酰-tRNA在16 S rRNA上的识别位点相同，这暗示IF1在翻译起始时可代替氨酰-tRNA暂时封闭核糖体的氨酰-tRNA接受位(A位)，起到协调30 S亚基功能的作用。另外，IF1是一个G蛋白，在核糖体亚基聚合时，具有活化GTP酶的作用。IF2有两种分子形式，IF2a与IF2b是同一个mRNA的翻译产物。IF2b的起始密码位于IF2a的+471核苷酸处。巯基是IF2活性所必需的，IF2磷酸化以后活性不变。IF2的功能是通过生成IF2·GTP·fMet-$tRNA_f^{Met}$三元复合物，在IF3存在下，使起始tRNA与核糖体小亚基结合。IF2与fMet-$tRNA_{Metf}$间作用的专一性非常严格，用于延伸的氨酰-tRNA甚至非甲酰化的Met-$tRNA_f^{Met}$均不能与IF2结合或结合得很不稳定。IF2具有很强的GTP酶活性，在肽链合成起始时催化GTP水解。IF3是具双功能的蛋白质，它至少有两种分子形式，IF3α和IF3β，后者比前者只少N端6个氨基酸残基。

IF3 对 mRNA 与核糖体的结合很严格，它与 16 S rRNA 相互作用位点在 700 与 790 环，840 茎与 1 500 连接区（图 9-9），其中 700 与 790 环与肽酰-tRNA 接受位点（P 位点）邻近。IF3 能通过促使未翻译的前导序列与 16 S rRNA 的 3′-端碱基配对，让核糖体识别天然 mRNA 上的特异的启动信号，又能刺激 fMet-tRNA$_f$ 与核糖体结合在 AUG 上。另外，IF3 能使 30 S 亚基形状发生细微的变化，以阻止其与 50 S 缔合。表 9-6 列出了 3 种起始因子以及蛋白质合成的延伸因子和终止因子的分子量和功能。

表 9-6　大肠杆菌的起始因子、延伸因子和终止因子的特性和功能

因子	分子量（kD）	特性和功能
超始因子		
IF1	9	促进核糖体的解离和 IF2 活性
IF2	100	由一个要求 GTP 的反应使 fMet-tRNA$_f$ 结合于核糖体的 P 位点
IF3	22	将 mRNA 结合于核糖体的小亚基，可能是促进非翻译的前导序列与 16 S rRNA 3′端的碱基配对
延伸因子		
EF-Tu	43	将氨酰-tRNA 结合于核糖体 A 位点
EF-Ts	30	重新生成 EF-Tu-GTP
EF-G	77	肽基-tRNA 和 mRNA 密码子和 A 位点移至 P 位点，此过程依赖 GTP
终止（释放）因子		
RF1	36	水解肽基-tRNA，要求 UAA 或 UAG 密码子
RF2	38	水解肽基-tRNA，要求 UAA 或 UGA 密码子
RF3	46	促进 RF1，RF2 活性

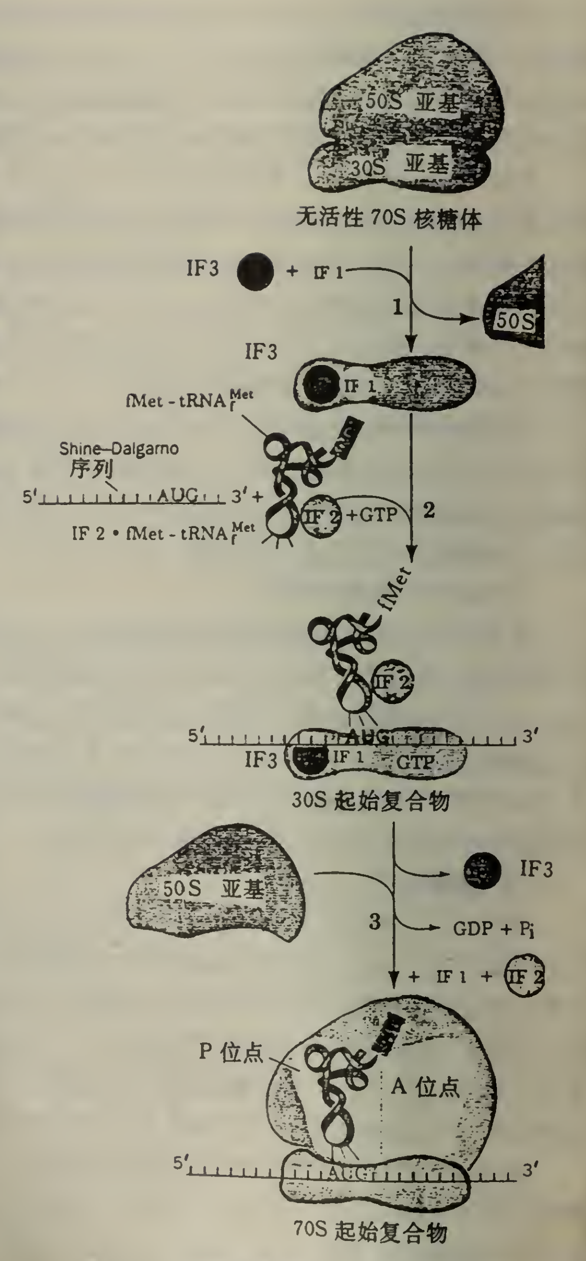


图 9-24　大肠杆菌蛋白质合成的起始过程

第二步，fMet-tRNA$_f^{Met}$ 和 mRNA 连接于 IF・30 S・GTP 聚集体上。在结合时，fMet-tRNA$_f^{Met}$ 与 IF2-GTP 复合物紧密接触。30 S 起始复合物一旦完全形成后，IF3 即释放出来，以后 50 S 参加进来，并引起 GTP 水解和释放其它两个起始因子（图 9-24），最后的复合物称为 70 S 起始复合物。

（五）核糖体上 tRNA 的结合位点

前文述及每一个核糖体含有 2 个 tRNA 结合位点的空腔，它们是 P（肽酰基）位点，A（氨酰基）位点。表 9-2 已指出 P 位点和 A 位点是由部分 30 S 亚基与部分 50 S 亚基形成的。由一

个特殊的 mRNA 密码子而使位点的形状对一给定的氨酰-tRNA 专一，就是说在该位点上的 mRNA 的表面使核糖体的空腔对任一给定的氨酰-tRNA 分子是专一的。

四、多肽链的延伸、移位循环

（一）肽链的延伸

从起始阶段形成的起始复合物可以接受第二个氨酰-tRNA，以形成蛋白质第一个肽键。在第二个氨酰-tRNA 进入 A 位点之后，便形成一个肽键，并产生出一个连接于第二个氨基酸的 tRNA 上的二肽。然后便发生移位，肽酰-tRNA 和与之结合的 mRNA 密码子协同转移至 P 位点。这个氨基酸加成过程一再重复，每次加上一个氨基酸，直至形成一条完整的多肽链（图 9-25）。

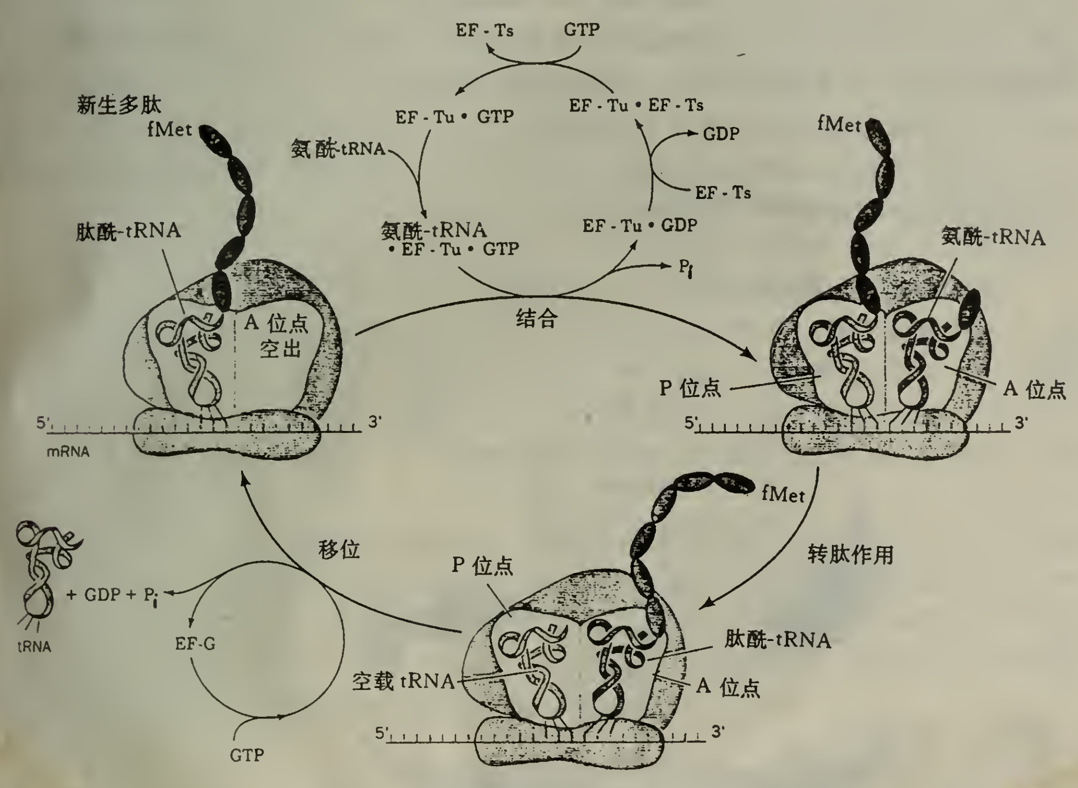


图 9-25　大肠杆菌蛋白质合成的延伸过程

肽链的延伸要求有延伸因子 EF-Tu 和 EF-Ts 参与。EF-Tu 的作用是帮助氨酰-tRNA 进入 A 位点。EF-Tu 先与GTP 结合，此时它处于活性状态。此 EF-TU · GTP 二元复合物（EF-TU · GTP complex）与氨酰-tRNA 形成 EF-Tu · GTP · 氨酰-tRNA 三元复合物。此三元复合物进入核糖体的 A 位点，而 P 位点则为 fMet-$tRNA_i^{Met}$ 或肽酰-tRNA 占据。以后 GTP 即被水解，生成无活性的 EF-Tu · GDP 被释放出。EF-Tu 是大肠杆菌内最丰富的蛋白质，每个细胞约含 100 000 个分子，与 tRNA 分子数约相等。

EF-Tu 是如何通过与 tRNA 受体末端的相互作用来促进反密码子在分子的另一端正确配对并不清楚，不过，在这个过程中 GTP 的水解却起了主要作用。如果三元复合物（即 GTP ·

EF-Tu·氨酰-tRNA）中的 tRNA 反密码子与结合在 A 位点的密码子完全不互补的话，该复合物将不会与核糖体相互作用。如果复合物的反密码子是接近于匹配，但不是十分正确，那么瞬间结合将会发生，GTP 也会水解，但是错配的 tRNA 将从核糖体上剔除。只有密码子和反密码子正确配对，在 GTP 水解后，氨酰-tRNA 的结合才会保持着。这样，凭借 EF-Tu 催化 GTP 的水解，似乎起着校正与阅读的机制，从而增加蛋白质合成中氨基酸排列的精确性。

最近 Powers 等提出，核糖体能觉察正确 tRNA 的结合，并能通过 16 S rRNA 的构象变化传递信息以引起 GTP 水解与 EF-Tu 释放。这种信息的传递步骤中有核糖体小亚基蛋白质 S 4，S 5 与 S 12，特别是围绕 16 S rRNA G530 环和 912～915 区的 RNA 结构域的参与。

1981 年 Rheinberger 等提出当有聚(U)存在时，有 3 个未经酰化的 tRNA 分子能和核糖体结合。2 个 tRNA 分子和 A 位与 P 位结合，第三个 tRNA 分子则和 E 位(exit，释放部位)结合。现在，已有许多证据表明大肠杆菌的 70 S 核糖体上除了 A 和 P 位外，还存在结合 tRNA 的第三个位点，即 E 位点。并证实移位前后的核糖体是处于两种不同构象状态。据此提出一种延伸循环变构 3 位点模型(图 9-26)。此模型认为核糖体有 A，P，E 3 个 tRNA 结合位点，E 位点与 P 位点相邻，偏向 mRNA 的 5′端，脱酰基-tRNA 离开 P 位点后转至 E 位点。这一结合使 A 位点的空间结构改变，利于新的氨酰-tRNA 的结合。当新的氨酰-tRNA 与 A 位结合后，使 E 位点结构改变，脱酰基-tRNA 即脱离 E 位点。

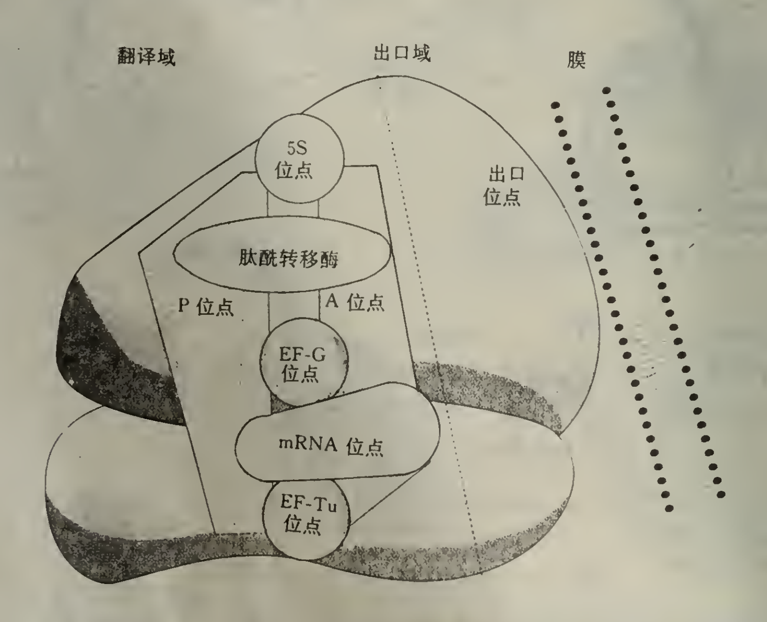


图 9-26 核糖体的 3 个位点及几个活性中心示意图，各活性中心的确切位置未定

EF-Tu 的再利用于携带另一个氨酰-tRNA 至核糖体 A 位点，则要求另一个因子 EF-Ts 的参与。EF-Ts 将 GDP 从 EF-Tu 置换下来，它本身则与 EF-Tu 形成一个复合物。当这个复合物遇到 GTP 时，便形成一个 GTP-EF-Tu 复合物，并放出 EF-Ts。EF-Tu-GTP 又可以与另一个氨酰-tRNA 结合而重复这个循环(图 9-25)。

移位前后，核糖体上的 2 个 tRNA 反密码子与 mRNA 密码子相互配对形成 6 bp，mRNA 折叠成 U 型转角，导致 A 与 E 位靠近，产生变构效应。最近(1996 年)，Agrawal 等用低温电子

显微镜(cryoelectron microscopy)技术直接观察到大肠杆菌的 70 S 核糖体上有 3 个 tRNA 结合,证实有 A,P,E 3 个位点存在。

另有实验证实,大肠杆菌核糖体的 E 位除为脱酰基 tRNA 离开核糖体提供出口外,E 位被正确的脱酰基 tRNA 占据时,A 位对不正确氨酰 tRNA 的结合可降低,而对正确氨酰tRNA 的结合几乎不受影响。相反,当 E 位存在不正确的脱酰基 tRNA 时,则可增加不正确的氨酰-tRNA 与 A 位结合。这些都意味着 E 位及 E 位上的密码子-反密码子相互作用,对蛋白质的生物合成及其翻译精确性也起着重要作用。

1989 年 Moazed 和 Noller 对 23 S rRNA 上的 tRNA 结合位点进行研究,提出了移位中间态假说,认为存在着杂合位点(hybrid site)。在肽链形成时,肽酰-tRNA 的 CCA 端从 P 位点移向 23 S rRNA 的 E 位点,其它部位仍位于 16 S rRNA 的 P 位点上,形成了一个复合位点P/E。同样,氨酰-tRNA 的 CCA 端由 A 位点移向 23 S 的 P 位点,其它部分也留在 16 S rRNA 的 A 位点上,即处于 A/P 位点。这样,移位是分两步进行。第一步是 tRNA 相对于 23 S rRNA 移动,形成杂合位点;第二步是完成移位后回到非杂合状态。这样,3 种 tRNA 就存在着 6 个结合位点。

(二)肽键的形成

肽键形成的酶促催化是核糖体大亚基的任务。催化肽键形成的酶叫肽酰转移酶(peptidyl-transferase)。大约 6 个 L 蛋白质和 23 S rRNA 对此酶的活性是必需的。另外,6 个蛋白和 5 S rRNA 也起很大作用。现认为它是把 2 个氨酰-tRNA 定于调准的位置,使之形成肽键。

肽键形成后,核糖体处于前移位状态。占领 P 位是去酰化 tRNA,A 位为肽酰-tRNA。移位过程需要延伸因子 EF-G 和 GTP。在核糖体存在下,EF-G 使 GTP 水解,故 EF-G 是一个依赖于核糖体的 GTP 酶。随着 GTP 水解,导致 EF-G 的释放。因此,除了起始需要 1 个 GTP 外,以后在生长中的多肽链上,每加上 1 个氨基酸就需要水解 2 个 GTP 分子。此外还需要 K^+离子和 Mg^{2+}离子参与。

1992 年 Noller 等发现催化肽键形成的不是肽酰转移酶,而是由 23 S rRNA 与其它临界生物学机能的 ribozyme 所催化。这个发现使人们对 RNA 在翻译中的功能做出新的评价。1994 年,Purohit 与 Stern 实验证实,解码是一种 RNA 的功能,是 rRNA 介导了 mRNA 与 tRNA 的相互作用的结果。他们认为密码子/反密子复合物的识别可能是 RNA 进化上的早期功能,而翻译的准确性可以看做是进化后期 RNA/蛋白质的功能。

五、多肽链合成的终止

蛋白质合成的终止需要两个条件:一个是应存在能特异地提出多肽链延伸应予停止的信号;另一个是有能解读链终止信号的蛋白质释放因子(RF)。由于多肽链延伸至其足够长度之后,其羟基端仍结合于 tRNA 接合体上。因此,终止应包括切除终端的 tRNA。当出现这个切除之后,新生的链便迅速离去,因为它主要是通过接合体 tRNA 结合在核糖体上的。

开始人们认为可能有肽链终止 tRNA 分子的存在,它们能专门地结合于终止密码子上,只是其 3′端处没有连接氨基酸,并且能以某种方式促进终端 tRNA 释放出tRNA。试验的结果却排除有这种 tRNA 分子的存在。终止密码子是被一种称为释放因子(release factor,RF)的蛋白质解读的。已知在大肠杆菌中,释放因子 RF1 能识别终止密码子 UAG 和 UAA,而 RF2

则识别UGA和UAA。每种释放因子先与GTP形成活性复合物，这个复合物再与终止密码子相结合，形成三元复合物，并且改变肽酰转移特异性。有释放因子存在时，肽酰转移酶催化肽基部分与水结合，而不是与游离氨基酰-tRNA结合。这样，肽酰转移酶便将P位上的肽基转移至水中（图9-27）。即水解P位上tRNA与肽链之间的键，随着便是新生肽链与最后一个脱酰基tRNA离开了核糖体。已知RF具有依赖核糖体的GTP酶活性，催化GTP水解，使RF与核糖体解离。在大肠杆菌中，还有一种释放因子，即RF3。它本身无识别终止密码子的功能，但却可以增加RF1和RF2的活性。现知核糖体结合与释放RF1和RF2都要受到RF3的刺激作用，后者可以与GTP和GDP相互作用。

肽链合成终止后，50 S核糖体亚基即释放。此时有一个被称为核糖体释放因子（ribosome releasing factor，RR）参与。在RR的作用下，核糖体与mRNA分离，同时携带最后一个脱酰基-tRNA与终止因子亦即脱落，RR起上述作用时，还必须有GTP和肽链延伸因子EF-G的存在。经过这一步骤，核糖体又可供重新合成另一条多肽链之用。

图9-27 肽酰转移酶反应

图示出因加入1个氨基酸而使多肽链的延伸，在链终止时，tRNA的3′-腺苷与氨基酸之间的酯键受到了取代另一个氨基酸的氨基的水的亲核攻击。

六、原核生物蛋白质合成中的某些抑制因素

原核生物的蛋白质生物合成为多种抗生素所抑制。某些抗生素可影响复制与转录而阻断蛋白质合成，但也有不少抗生素则是直接妨碍翻译过程。

（一）嘌呤霉素

嘌呤霉素的结构很类似氨酰-tRNA，故能与后者相竞争，作为转肽反应中氨酰异常复合物，从而阻断蛋白质的生物合成。当生长着的肽链（或甲酰甲硫氨酸）被转移到嘌呤霉素的氨基上时，新生的肽酰-嘌呤霉素会从核糖体上脱落下来，从而终止翻译。肽链合成过早终止，而且

截短的肽链的羧基端连有1分子的嘌呤霉素。

（二）链霉素

链霉素能与30 S 亚基结合。此30 S 链霉素复合物是一种效率很低的并且不稳定的起始复合物，很容易解离而终止翻译。链霉素结合在30 S 亚基上时，能改变氨酰-tRNA 在 A 位点上与其对应的密码子配对的精确性与效率。

（三）四环素

四环素能阻断氨酰-tRNA 进入 A 位点，从而抑制肽链的延伸。

（四）氯霉素

氯霉素能抑制核糖体中的50 S 大亚基的肽酰转移酶的活性，抑制肽链的延伸。

（五）红霉素

红霉素与50 S 亚基结合，抑制肽酰转移酶，妨碍移位，因而将肽酰-tRNA“冻结”在 A 位上。

由于细菌核糖体与哺乳动物线粒体核糖体相似，因此，能抑制细菌核糖体的抗生素也能抑制哺乳动物线粒体的核糖体循环，但它们并不与哺乳类通常核糖体结合，因此，它们对哺乳动物蛋白质合成影响不大。四环素类（土霉素、金霉素、四环素）虽然对细菌与哺乳动物两种核糖体都能结合，但因为它进入细菌体比进入哺乳动物的细胞器要容易些，所以，对细菌有选择作用。

第五节　GTP 在蛋白质合成中的作用

一、GTP 在核糖体反应中的作用

GTP 和 ATP 一样是高能分子。一般来说，当这样的分子水解时，水解的自由能用来驱动需能的反应进行。但在蛋白质合成中则似乎并非如此。考察反应顺序表明，GTP 的作用是使各种翻译因子（IF2，EF-Tu，EF-G）与 tRNA 或核糖体易于以非共价键结合。这些结合要求 GTP 的存在。此外，GTP 水解成 GDP 和 Pi 的反应则是释放出结合因子的信号。比较了游离因子和因子-GTP 复合物的结构表明，当 GTP 结合时，诱导翻译因子发生形状变化，而这正是它们结合于核糖体所必需的。因此，GTP 是一种变构因子，即可以通过结合诱使大分子形状发生变化的小分子，当 GTP 与之结合时，大分子为活性构象；当 GTP 水解或除去时，大分子回复其无活性形式，结果是将相应的因子从核糖体表面排出。

重组试验证明，在50 S 亚基上的4个 L7/L12 蛋白拷贝是与 GTP 水解密切相关的。这些核糖体组分对各种蛋白质合成因子的结合是必需的，当它们起作用时，则利用贮存于 GTP 中的能量。但是，在没有 GTP 水解和延伸因子存在时，也能进行蛋白质的合成，但速率很慢。这样，核糖体本身必须具备本来就具备的结合氨酰-tRNA 和移位的能力，各种因子和 GTP 的参与则是增加这个速率。

除上述功能之外，GTP 的另一个极其重要的功能是使翻译准确。不少证据证明，在氨酰-tRNA 结合后有一个校正步骤。即氨酰-tRNA 最初在 A 位点与 EF-Tu-GTP 结合成复合物时，起始的 mRNA-tRNA 碱基配对相互作用并没有足够的特异性可用以说明蛋白质合成的低的误差率（约 10^{-4}）。凭借摒弃不正确的氨酰-tRNA 以进行校正，便可以降低其误差率。这就要

消耗一个氨酰-tRNA-EF-Tu-GTP 复合物(包括 ATP→GDP 的转变),而多肽链并不延伸。这样,校正步骤便要求有添加的能量输入。

二、鸟苷 5′多磷酸 3′多磷酸化合物(p)ppGpp 的生成

如果偶然地一个无载荷的 tRNA 占据 A 位点时,不仅多肽生长暂时停止,并且在受影响的核糖体上发生一个空载反应(idling reaction),这个反应是用 ATP 做焦磷酸的供体,将 pp5′-G-3′-OH(GDP)转变为稀有的鸟嘌呤核苷酸 pp5′-G-3′pp。现知只有当无载荷的 tRNA 正确地与暴露于 A 位点的 mRNA 密码子配对时,才会促进 pp5′-G-3′pp 的生成。pp5′-G-3′pp 简写为 ppGpp。1969 年 Cashel 发现此化合物时,称其为魔点Ⅰ(magic spot Ⅰ,MSⅠ),将 pppGpp 称为魔点Ⅱ(MSⅡ)。其产生的反应如下:

$$\text{(p)ppG}+\text{pppA}\xrightarrow[\text{SF}]{\text{核糖体·mRNA·tRNA}}\text{(p)ppGpp}+\text{pA}$$

合成(p)ppGpp 时需要核糖体的 2 个亚基、mRNA 以及对核糖体 A 位上密码专一的游离(即不带氨基酸)的 tRNA。它们是在蛋白质合成的"空步"上生成的。反应式中的 SF(stringent factor)是"严格控制因子",它是 *relA* 基因产物,是联系在核糖体上的一种蛋白质,分子量为 77 kD。现认为它即是催化 pppGpp 和 ppGpp 合成的酶。

实验得知,细菌正常生长时只生成极少的 ppGpp,当细菌缺乏必需的氨基酸时,细菌体内发生一系列的变化来应付不利的环境。这些变化主要表现为对体内各种代谢的"严格控制"。

tRNA 和 rRNA 的合成被抑制,延伸因子 EF-Tu,EF-Ts,EF-G 以及其它一些核糖体蛋白质如 S3,S4,L1,L2,L3,L4,L7,L10,L11,L12 等的合成也被抑制,细菌的不少代谢如糖原水解、糖的氧化、磷脂合成、脂肪合成、细胞壁多糖的形成、膜的转运、核苷酸的合成、多胺的吸收和合成等都被抑制;有少数代谢如糖原合成和氨基酸合成以及蛋白质的水解却被促进。许多实验表明 ppGpp 和 pppGpp 是细菌"严格控制"反应的多效应的信号分子,这样就保证细胞不会制造多于其需要的核糖体而造成浪费。这种在氨基酸缺乏条件下对 rRNA 和 tRNA 合成的专一性的阻塞称为严格应答(stringent response)。可是它的存在不是强制性的。因为有些松弛型突变型(relaxed mutant)即使在氨基酸缺乏时,也继续合成 rRNA 和 tRNA 链,这是因为松弛型突变型细胞缺少 SF 因子的缘故。其它松弛型突变型则是 50 S 核糖体蛋白 L11 发生了变化。这种"严格控制"系统在原核生物中具有普遍性,真核生物中有无该控制系统,目前尚无定论。

第六节　真核生物的蛋白质的生物合成

真核生物蛋白质合成与原核生物两者相比,密码相同,各种组分相似,亦有核糖体、tRNA 及各种蛋白质因子。总的合成途径也相似,有起始、延伸及终止阶段,但也有不同之处:

(1)真核生物的蛋白质合成与 mRNA 的转录生成不偶联。mRNA 在细胞核内以前体形式合成,合成后需要经加工修饰才成熟为 mRNA,从细胞核内输入细胞质然后投入蛋白质合成过程。

(2)真核生物蛋白质合成机构比原核生物复杂,起始步骤涉及起始因子众多,过程复杂。

(3)真核生物蛋白质合成的调控复杂。

(4)真核生物与原核生物的蛋白质合成可为不同的抑制剂所抑制。

一、真核生物蛋白质合成的起始

真核细胞的 mRNA 在 5′-末端“戴”有一个三磷酸 7-甲基鸟苷(m^7GpppN)的帽子，3′-末端常见有一个由 50～200 个残基构成的多聚腺苷酸(polyA)的尾(图 9-21)。在细胞核内的 mRNA前体中常含有插入序列，它相当于基因中的内含子。输入细胞质前，mRNA 还与很多蛋白质结合成核蛋白，即信使核糖核蛋白。现知戴“帽”与加“尾”在细胞核内与细胞质中都可以进行。帽的结构在蛋白质合成的起始过程具有重要作用。它能增强翻译。没有甲基化的帽子(GpppN—)，以及用化学方法或酶学方法脱去帽子的 mRNA，其翻译活力显著下降。帽子还会保护 mRNA 不受核酸外切酶的侵袭，使 mRNA 相对稳定。近年来对尾的研究表明，poly(A)尾的功能不仅显示在 mRNA 由核向细胞质转运的过程中，而且它对 mRNA 的稳定性以及翻译效率均有调控作用。现知 poly(A)尾对 mRNA 分子的稳定作用需要有 poly(A)结合蛋白(PABP)的参与，而对翻译的促进还与 poly(A)尾的长度有关。

与原核生物不同，真核生物的起始 $tRNA_i^{Met}$ 所携带的是甲硫氨酸。虽然在体外可被由大肠杆菌所提取的甲硫氨酸 tRNA 转甲酰基酶催化而甲酰化，但在体内携有甲硫氨酸的 tRNA 分子上，其甲硫氨酰部分并未甲酰化，即可在肽链合成中起始反应。

参与翻译起始反应的起始因子已发现有十几种，其特性与功能见表 9-7。从图 9-28 可见真核生物蛋白质生物合成的起始过程。这个过程可以分为 3 个步骤：

(一)43 S 前起始复合物的形成

在起始因子 eIF-3 的作用下，80 S 核糖体解聚为 40 S 和 60 S 亚基。eIF-3 并有防止这 2 个亚基再结合的作用。另一因子 eIF-4C(eIF-1A)也有助于这个解聚。起始因子 eIF-2 与 GTP 形成稳定复合物，后者与 Met-$tRNA_i^{Met}$ 形成三元复合物，再与 40 S 亚基形成 43 S 前起始复合物(图 9-28 a)。

(二)mRNA 的结合

在起始因子 eIF-4A，eIF-4B，eIF-4E 和 ATP 的参与下，43 S 与 mRNA 结合。eIF-4A 有使 mRNA 二级结构解旋的作用。eIF-4B 则有结合 mRNA 并识别起始密码子 AUG 的作用。上述的 eIF-3 也参与 40 S 三元复合物与 mRNA 结合的作用。eIF-4E 或称“帽子结合蛋白Ⅰ”(cap binding protein，CBPⅠ)，起与 mRNA 帽子结合的作用。另有 eIF-4F，又称 CBPⅡ，实际是包括 CBPⅠ和 eIF-4A 和一种分子量为 220 kD 的蛋白(P220)。此外还有 eIF-6，与 60 S 亚基结合使核糖体保持在解聚状态。

由 40 S 亚基、Met-$tRNA_i^{Met}$ 和一些起始因子组成的前起始复合物在 mRNA 的 5′-帽子处或其附近与之结合，然后沿着 mRNA 滑动，直至遇上第一个 AUG 密码子(图 9-28 b_1)。这个过程由 CBP 促进，并消耗 ATP，使 mRNA 的 5′端二级结构解旋，使它呈线状穿过 40 S 亚基颈部的通道。CBP 是在 mRNA 的 5′端识别帽子的，以后 eIF-4A 和 eIF-4B 也参与沿着 mRNA 的解旋。43 S 前起始复合物与 mRNA 的结合，是在帽子结构下游 50～100 个核苷酸范围内。据 Kozak 等的研究，大多数起始密码子的合适‘上下文’为 CCACCAUGG。在 43 S 前起始复合物沿 mRNA 向 3′端方向移动时，遇到合适的‘上下文’，即停止移动。起始密码子 AUG 的识别可能是通过与 tRNA 上的反密码子配对的作用。eIF-2 也参与这个识别的作用。以后，便形成 48 S 前起始复合物(图 9-28 b_2)。

表 9-7　真核生物蛋白质合成的起始因子

因　子	分子量 (kD)	亚基 (kD)	功能	旧名称
eIF-1	15	15	促使 43 S 和 48 S 起始复合物的形成	
eIF-1A	18	18	使核糖体解聚	eIF-4C
eIF-2	130	α36	依赖 GTP 使 Met-tRNA$_i$ 结合到小亚基上	
		β38		
		γ55		
eIF-2B	272	α 26	eIF-2・GDP	GEF，anti-HRI
		β 39	转变为 eIF-2・GTP	
		γ58		
		δ67		
		ε82		
eIF-3	550	α35	与 40 S 亚基结合，保持解聚 80 S 核糖体，结合 mRNA 于 43 S 起始复合物前体	
		β36		
		γ40		
		δ44		
		ε47		
		ζ66		
		η115		
		θ170		
eIF-3A	25	25	使核糖体解聚并与 60 S 亚单位结合	
eIF-4A/eIF-4β	44		具有 RNA 依赖的 ATP 酶、解旋酶活性，使 mRNA 结合到 43 Sn 上	
eIF-4B	80	80	使 mRNA 结合到 43 S 复合物上，识别 AUG	
eIF-4E(eIF-4α)	25	25	帽子结合蛋白	CBP-Ⅰ
eIF-4F	270	α25	识别帽子结构，具有 RNA 依赖的 ATP 酶、解旋酶活性，使 mRNA 结合到 43 S 前起始复合物上	CBP-Ⅱ
		β44		
		γ220		
eIF-5	60	60	使 eIF-2 键合的 GTP 水解释放 eIF-2 和 eIF-3 以使释放 eIF-2 和 eIF-3 以便 60 S 亚基连接上	
eIF-5A	15	15	使甲硫氨酰基嘌呤霉素形成	eIF-4D
eIF-6	24		与 60 S 亚基结合，保持亚基解聚	

注：43 Sn 为 43 S 天然起始复合物；CBP：帽子结合蛋白；GEF：GTP-GDP 交换因子。

（三）80 S 起始复合物的形成

在形成 48 S 前起始复合物之后，再与核糖体的 60 S 大亚基结合，最后便形成 80 S 起始复合物。在另一起始因子 eIF-5 的作用下与 eIF-2 键合的 GTP 被水解，并释放出 eIF-2-GDP，Pi 和 eIF-3。其它起始因子也释放出来，最后形成由 40 S 亚基・Met-tRNA$_i^{Met}$，mRNA・60 S 亚基组成的起始复合物（图 9-28 c）。释放出的起始因子可以再用于下一轮的起始复合物的形成而循环利用。在起始因子中，Met-tRNA$_i^{Met}$ 是在核糖体的 P 位点，以后其中甲硫氨酸与另一氨酰-tRNA形成二肽酰-tRNA。

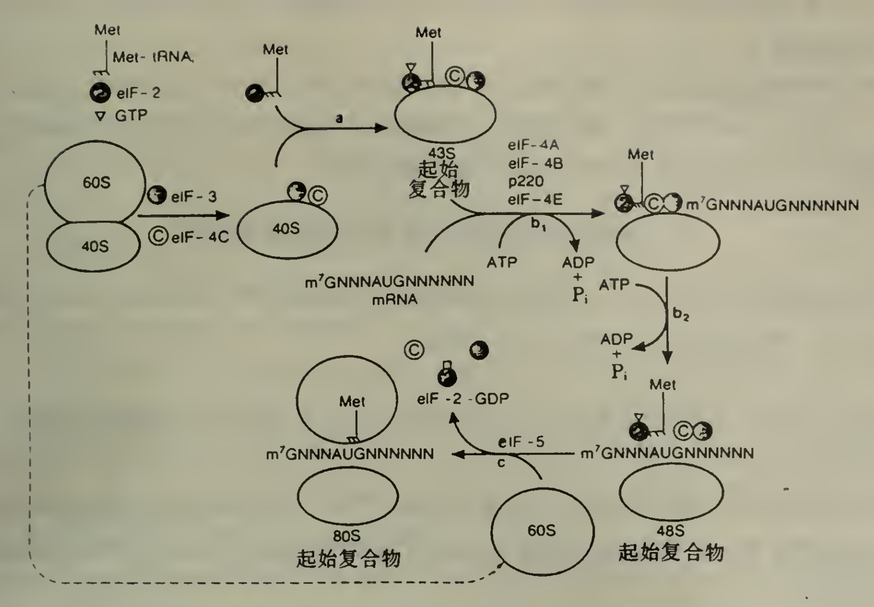


图 9-28　哺乳动物(真核生物)蛋白质合成的起始复合物形成

二、肽链的延伸与终止

(一)肽链的延伸

真核生物的肽链延伸与原核生物相似，只是延伸因子 EF-Tu 和 EF-Ts 被 eEF-1 取代，而 EF-G 则被 eEF-2 取代(图 9-25)。在真菌中，还要求第三种因子，即 eEF-3 的参与，以维持其翻译的准确性。

延伸因子 eEF-1 是个多聚体蛋白质，大多数由 $\alpha,\beta,\gamma,\delta$4 个亚基组成。eEF-1$\alpha$ 的分子量是 50 kD，作用与 EF-Tu 相似，与 GTP 和氨酰-tRNA 形成复合物，并把氨酰-tRNA 传递给核糖体。在每一轮循环中，GTP 在 eEF-1α 从核糖体上释放之前被水解。eEF-1β 具鸟苷酸交换活性，作用与 EF-Ts 相似，eEF-1γ 常与 eEF-1β 形成复合物，增加后者的 GDP-GTP 交换功能。在脊椎动物中还有 eEF-1δ，它与 eEF-1β 具同源性。这样，eEF-1$\alpha\beta\gamma\delta$ 复合物就包含有两个蛋白质交换因子，而在原核生物中只有 EF-Ts 一个交换因子。

延伸因子 eEF-2 是个单体蛋白。分子量约 100 kD。相当于原核生物中的 EF-G，催化 GTP 水解，使氨酰-tRNA 从 A 位转移至 P 位。Nilsson 报道，eEF-2 可以与核糖体形成高亲和力与低亲和力两种形式的复合物，前者相当于转位前状态，后者相当于转位后状态，并且在后者中，其 GTP 酶活性被激活。eEF-2 可与 GDP 形成稳定的二元复合物，与 GDP 的结合比与 GTP 的结合高 10 倍。eEF-2 近氨基端的 Thr_{56} · Thr_{58}残基被磷酸化后，会使延伸速率降低，从而达到调控目的。

延伸因子 eEF-3 是在真菌中发现的。是一条 120～125 kD 的多肽链，可结合 GTP，亦能水解 GTP 与 ATP，eEF-3 在翻译的校正阅读方面起重要作用。Uritani 推断，在酵母核糖体上存在一个引入位点 I (introductory site)。由 eEF-3 介导，使 eEF-1α · GTP · 氨酰-tRNA 三元复合物先结合到该位点上，然后再以密码依赖的方式进入 A 位点。这样使正确的 tRNA 进入得

到保证。eEF-3 的 ATP 酶活性可能与氨酰-tRNA 在核糖体上位点之间移动有关。

(二)肽链的终止

真核生物的肽链合成的终止仅涉及一个释放因子 eRF。eRF 分子量约 115 kD。它可识别 3 种密码子 UAA,UAG,UGA。eRF 在活化了肽酰转移酶释放新生的肽链后,即从核糖体上解离。解离要求 GTP 的水解。故终止肽链合成是耗能的。

三、真核生物蛋白质生物合成的特异抑制剂

真核生物的蛋白质合成有一些特异的抑制剂,如白喉毒素(diptheria toxin)、志贺氏毒素(shigella toxin)、蓖麻蛋白(ricin)等,在此不详述。

第七节　蛋白质折叠与蛋白质生物合成中多肽链的修饰

上文讨论的 mRNA 翻译得到的多肽链是初级产物。由它们转变成相应的天然蛋白质,包括从末端或内部位置去除某些氨基酸残基,某些氨基酸侧链的改变,以及多肽链折叠成适当的二级与三级结构。

一、蛋白质的折叠

蛋白质折叠是指从多肽氨基酸序列形成具有正确三维空间结构的蛋白质的过程。肽链的折叠从核糖体出现新生的多肽链即可开始。

(一)体外蛋白质的复性

60 多年前,吴宪教授提出了蛋白质变性(denaturation)学说,即:“天然蛋白质之分子,因环境种种之关系,从有秩序而坚密之构造,变为无秩序散漫之构造,是为变性作用。若数分子相撞而缠结,则为凝固作用。蛋白质变性之种种事实,均可以此说解释之”。可见,失去活性的蛋白质是处于伸展状态。如果一旦去除变性因素,恢复到“有秩序而坚密之构造”,称为复性(renaturation)。60 年代初,Anfinsen 的实验表明,失去活性的胰脏核糖核酸酶 A,在适当条件下,能够再折叠成呈活性的酶。与此同时(1961 年),杜雨苍等根据胰岛素 A,B 链拆合的实验就指出:亚基间 S—S 的形成有高度选择性,不是随机的。只要有了一级结构,就能形成天然高级结构。这些研究成果说明,氨基酸序列是决定蛋白质空间构象的最基本因素。越来越多的事实说明,一维多肽序列向三维蛋白空间结构信息传递,必然存在某些迄今未被揭示的内在规律。有人亦称之为“第二密码”或“折叠密码”。把其规律揭示出来正是当今分子生物学研究热点之一。

(二)体外蛋白质折叠的机制

变性的蛋白质能复性的现象,暗示在适合的条件下,一级结构具有自动恢复破坏了天然结构的能力。近年来的研究表明,伸展肽链(U)经早期的变化进入中间态(I),然后由中间态过渡到最终的天然态(N)。

$$\mathrm{U} \xrightarrow{\text{快}} \mathrm{I} \xrightarrow{\text{慢}} \mathrm{N}$$

伸展的肽链从什么地方开始折叠?这问题目前仍未有统一的意见。有人认为:伸展的肽链上已有某些扭曲、转角或链的不同部位之间的相互作用的区域。这些亚结构可能是不稳定的,但在折叠开始阶段它们可能起到“种子”作用,围绕着它们形成较稳定紧密的区域,然后形成稳

定的二级结构。计算机模拟认为，折叠可能起始于转角。转角属于稳定的二级结构。这些转角在折叠过程中一直存在，直到最后被编入三维结构中。

另一观点认为折叠起于由疏水作用而引起的坍塌，称为疏水坍塌(hydrophobic collapse)。结果是疏水侧链集中埋藏于分子的内部，而极性侧链则裸露于与水接触的分子表面。

也有人认为，折叠起始于共价键的相互作用的形成。Oas 与 Kim 合成了牛胰胰蛋白酶抑制剂的两个片段。一个为 16 肽，相当于氨基酸顺序的 43～58；另一为 14 肽，相当于第 20～33 氨基酸。它们分别在第 30 与 51 位各有一个 Cys 残基。在还原条件下，两个肽都是伸展的。但当两个肽的 Cys 30 和 Cys 51 以二硫键连接起来时，各自表现出天然分子的构象。说明二硫键形成使两个肽段都获得稳定的二级结构。

Ptitsyn 研究了牛或人 α-乳清蛋白在酸性、热变性、中等浓度盐酸胍溶液条件下的各种物理性质。发现在这些条件下，形成了一种彼此类似但与天然 α-乳清蛋白不同的新的物理性状，这种形式被称为“熔球态”(melton-globule state)。熔球态是目前比较公认的折叠过程中的中间态。与天然分子相比，熔球态波动性增加，不存在特定的牢固安排的侧链基团，但其结构紧密，具有含量很高的类似天然分子的二级结构。

概括起来，体外蛋白质的折叠可能是始于疏水坍塌，或始于转角，或始于共价键相互作用(如二硫键的形成)。在折叠早期，可能这三种方式联合起作用。之后，可能沿着有限的多途径形成中间态(熔球态)。这个过程是快速的，一般在毫秒范围内完成。最后再由中间态进入天然态，此过程比较慢，是折叠反应的限速步骤。

(三)体内蛋白质的折叠

体内蛋白质的折叠与体外的很不相同。体外蛋白质折叠的环境单纯，折叠从完整的肽链开始；体内的则环境复杂，有很多其它因子与蛋白质参与。从效率上看，体内的效率高，而体外的效率低。

与较早期的认为蛋白质的折叠是发生在完整的多肽链合成之后的观念相反，邹承鲁提出了新生肽折叠的假说。他认为体内蛋白质的折叠可能存在着与肽链合成同步进行的过程(co-translational)。随着肽链延伸，其空间构象不断调整，合成终结后再进一步形成天然态(post-translation)。实验证实，肽链合成尚未结束的 β-半乳糖苷酶已具有免疫原性和酶的活性，表明酶蛋白的 N 端部分就有天然态的结构与功能。这一事实有力地支持了邹氏模型。体内蛋白质的折叠远比体外环境复杂，有很多其它因子与蛋白质存在，正是由于它们的参与使错误的折叠与聚集作用大大降低，从而提高了效率。实验证实，体内新合成的野生型蛋白质肽链，有 95% 以上能转变为天然的三维结构。

近来的研究表明，至少有两类蛋白质参与体内的折叠过程，称为助折叠蛋白(folding helper)：

1. 酶　如蛋白质二硫键异构酶(protein disulfide isomerase，PDI)与肽酰脯氨酰顺反异构酶(peptidyl prolyl cis/trans isomerase，PPI)。前者能加速蛋白质正确二硫键的形成，后者催化肽-脯氨酰之间的肽键的旋转反应。这些反应与新生肽链的折叠密切相关，加速蛋白质折叠过程。

脯氨酸吡咯环上的亚氨基在形成肽键时，可以形成顺式或反式结构；反式结构是较为稳定的。在蛋白质中，也有一部分脯氨酸亚氨基的肽键是顺式的。蛋白质变性后，便全部异构化为反式。在复性时，必须再异构化为顺式，才能恢复蛋白质原来的空间结构。在体外，这一反应是一个较慢的反应。在体内，在肽酰脯氨酰顺反异构酶的催化下，使这一顺反异构作用快速进行。

二硫键存在于大多数分泌蛋白及膜蛋白中。在翻译产物形成的早期，都是以还原的巯基形式存在，然后再氧化生成二硫键。在体外，蛋白质中的巯基为空气氧化生成二硫键反应很慢。在体内，在蛋白质二硫键异构酶的催化下，新生肽键的巯基氧化生成二硫键可以加速进行。

2. 分子伴侣　分子伴侣最早是由 Laskey 等人提出来的。他们发现组蛋白必须先与核内的核质素(nucleo-plasmin)结合后，才能与 DNA 组成核小体。他们将这个核质素称为分子伴侣(molecular chaperone)。后来 Ellis 等，把一类在细胞内帮助新生肽链正确组装，成为成熟蛋白质，而本身却不是最终功能蛋白质分子的组成成分的分子，都称为分子伴侣。分子伴侣促进一个反应的进行，而本身却不出现于最终产物，具有类似于酶的特征，但它又与酶有很大差异。它对靶蛋白质专一性不高。同一个分子伴侣可以促进多种氨基酸序列完全不同的多肽链折叠成为空间结构、性质和功能都不同的蛋白质。而且它的“催化”效率很低。分子伴侣的作用有时也只是阻止肽链的错误折叠，而不是促进其正确折叠。

成熟的功能蛋白质的空间结构是疏水性基团在分子内部，亲水性基团在分子的外部，使蛋白质分子具有水溶性。在新生肽链合成过程中，先合成的部分通常会含有疏水性的侧链，肽链开始折叠时，可能在合成和折叠的早期已经形成一定的疏水性结构。在整个分子合成完毕之前，在先形成的疏水性结构之间可能发生错误的结合，产生肽链的错误折叠和部分疏水性结构瞬间暴露在分子外部而引起沉淀。分子伴侣可能识别多肽链中某些先合成的并部分折叠成的疏水结构并与之结合，使疏水表面间不会形成错误的相互作用，从而防止多肽链的错误折叠。分子伴侣的作用在促进靶蛋白质正确折叠或阻止其错误折叠的作用后，便从靶蛋白质上脱离。

目前已鉴定很多分子伴侣蛋白质，其中被研究得最多的有两个蛋白质家族，即胁迫-70(stress-70)家族与分子伴侣(chaperonin)家族，这些蛋白质广泛存在于原核与真核生物细胞中。

(1)胁迫-70 家族　本家族成员的分子量约为 70 kD。由两个结构域组成。不同来源的胁迫-70 蛋白的 N 端的结构域高度保守，具有 ATP 酶活性。C 端的结构域保守性较差，是与多肽链结合部位。胁迫-70 蛋白在细胞内含量丰富。研究较多的是其家族成员之一的热休克蛋白(heat shock protein，Hsp)70，Hsp70。现在已知，热休克蛋白除参与蛋白质的折叠外，在蛋白质的组装(assembly)、跨膜、分泌与降解都有它们的参与。

Flynn 等用一套随机氨基酸顺序但长度有限的合成的多肽与胁迫-70 蛋白之一的重链结合蛋白 Bip 相互作用，观察到由四肽至七肽之间，Bip 与其结合的能力是随着肽链的长度延伸而增加，并且刺激 ATP 水解的能力也随之增加，但肽链再延长则上述能力不见增长，说明结合部位只能容纳七肽。同时还发现，Bip 偏爱与脂肪侧链的多肽相结合，预示这种结合可能是疏水相互作用。现知新生肽链合成过程中，首先合成的部分通常也会有疏水侧链。肽链开始折叠时，可能在合成过程中已经形成一定的疏水结构。这样，一个新生的伸展的肽链就可能通过疏水作用与若干个胁迫-70 蛋白相结合。

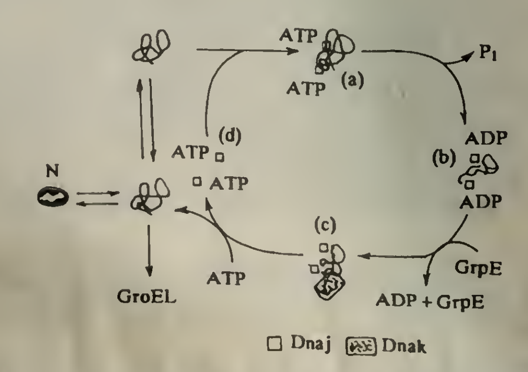


图 9-29　DnaK，DnaJ 和 GrpE 在蛋白质折叠中的反应循环

在大肠杆菌中的 Hsp70 称为 DnaK，它的功能受两个协同因子热休克蛋白 DnaJ(一种温

度敏感性蛋白)与GrpE(一种富甘氨酸蛋白)的调节。伸展的肽链在DanJ的介导下与DnaK结合,从而形成多肽·DnaK·DnaJ三联体复合物(图9-29 a)。DnaJ能使DnaK上的ATP水解(图9-29 b),另一方面,DnaJ又能使DnaK-ADP稳定,而稳定的DnaK-ADP对伸展的肽链具有很高的亲和性。此时被结合其上的多肽链不会进一步折叠,也不会错误折叠或聚集成不溶形式。之后,在GrpE作用下,ADP从DnaK上释放(图9-29 c),结合其上的多肽链也随之脱落,DnaK恢复结合ATP的能力。释放的肽链或被折叠成天然态,或随着ATP的水解再次形成多肽·DnaK·DnaJ三联体复合物或被转移给另外的监护分子。

(2)分子伴侣家族　监护蛋白有两种氨基酸顺序相关的类型。大的叫分子伴侣-60(cpn-60),是一个寡聚体,由14个相同亚基形成双层饼状,亚基分子量约为60 kD。具有多个肽链的结合位点。在大肠杆菌中称为GroEL的是cpn-60家族成员之一。小的叫分子伴侣-10(cpn-10),也是以寡聚体形式存在。由7个相同亚基组成单层环状。亚基分子量为10 kD。在大肠杆菌中称为GroES的是其cpn-10家族成员之一。

在生理条件下,GroEL与GroES形成复合物,形如空圆筒状。非天然底物蛋白就结合在GroEL空圆筒上的开放空穴上,像是"帽子"一样。然后通过一系列ATP的水解,使结合的蛋白质被释放到中空的筒中,并使之折叠,如折叠完成就被释放,如未完成折叠就被再次结合,进行第二次循环。

已知分子伴侣所识别的靶蛋白是它们的部分折叠的非天然态,但是分子伴侣怎样去识别它们,还不清楚;另外,GroEL水解ATP,可以加速结合的底物蛋白质的释放,并且,即使GroES缺乏时,ATP也可使底物蛋白质成功地折叠。可是,ATP水解所释放的能量怎样与折叠相偶联仍知之甚少。

应该指出,就蛋白质的天然态、伸展态与中间态而言,它们之间存在能量上的差异,尽管其差异较小。已知一个体系处于能量最低态时是最稳定的,但是,天然态的蛋白质并非是能量的最低状态,因为蛋白质在执行生物学功能时,就要求其构象具有一定的柔性(flexibility),例如,酶的活性中心即是如此。因此,蛋白质的折叠是与其体系的自由能和折叠的途径有关,即受制于体系的热力学,也与其动力学有关。

二、蛋白质的修饰

由核糖体合成的多肽链除需要折叠成正确的空间结构外,还必须进行修饰,才能成为具有生理功能的蛋白质。修饰可以在肽链折叠之前、或折叠期间、或折叠之后进行,也可以在肽链延伸期间或在终止之后进行。有些修饰对多肽链的正确折叠是重要的。有时,修饰也与合成的蛋白质在细胞内的转移或分泌(见下节)有关。

(一)末端氨基的脱甲酰化和N端甲硫氨酸的切除

大肠杆菌等细菌的多肽链合成的起始氨基酸是甲酰甲硫氨酸,在翻译后这个甲酰基被肽脱甲酰基酶(peptide deformylase)除去,甲硫氨酸通常也被氨肽酶切除。

真核生物的多肽链合成的起始氨基酸是甲硫氨酸。这个甲硫氨酸也只暂时存留,在肽链延伸一定长度后便被切除,例如,珠蛋白在合成了约30个氨基酸残基后便将起始的甲硫氨酸切除。

(二)多肽链的水解断裂

新生多肽链常发生水解断裂,生成较短的肽链。例如,胰岛素的合成,先生成较大的前体,即前胰岛素原(preproinsulin),然后水解断去一段由－1～－24位氨基酸的N端信号序列(称为前

肽 pre-peptide)，并形成二硫键，生成胰岛素原(proinsulin)。胰岛素原再由一肽链内切酶在两处切去两对碱性氨基酸，并由肽链外切酶再切去一段连接的肽链，最后生成胰岛素的两条以二硫键连接的 A，B 链。一些酶也是先形成无活性的酶原，经切去一段肽链后才变为有活性的酶。

（三）氨基酸侧链的修饰

1. 二硫键形成　上文已提及，在多肽链中的 Cys 残基，可在蛋白质二硫键(disulfide isomerase)异构酶的作用下形成二硫键。这个反应可在翻译过程中或在翻译后进行。

2. 羟化作用　例如胶原蛋白在合成后，其中的某些脯氨酸和赖氨酸残基发生羟化。在 X-Pro-Gly(X 代表除 Gly 外的任何氨基酸)序列中的脯氨酸羟化为 4-羟脯氨酸；也可生成 3-羟脯氨酸，但较少。脯氨酸的羟化有助于胶原蛋白螺旋的稳定。有一些赖氨酸也羟化为 5-羟赖氨酸，以后再糖基化(见下文)。其它蛋白质如核心专一凝集素(core-specific lectin)也含有羟脯氨酸和羟赖氨酸。

3. 氨基酸残基的交联　在一些结构蛋白如胶原蛋白和弹性蛋白中，多肽链常发生交联(cross-linking)，除形成二硫键外，也包括其它氨基酸的交联。例如，胶原蛋白中的一些特殊的赖氨酸和羟赖氨酸氧化为醛衍生物 ε-醛基赖氨酸(allysine)，以后 2 个发生醛醇缩合形成醛醇桥(aldol bridge)，或与另一赖氨酸缩合并还原，形成仲胺桥(图 9-30)；另一例子是谷氨酰胺的酰胺基与赖氨酸侧链中的氨基缩合。这个反应是在谷氨酰胺酶(glutaminase)催化下进行的，它在血液凝固时血纤维蛋白单体的交联中起重要作用。

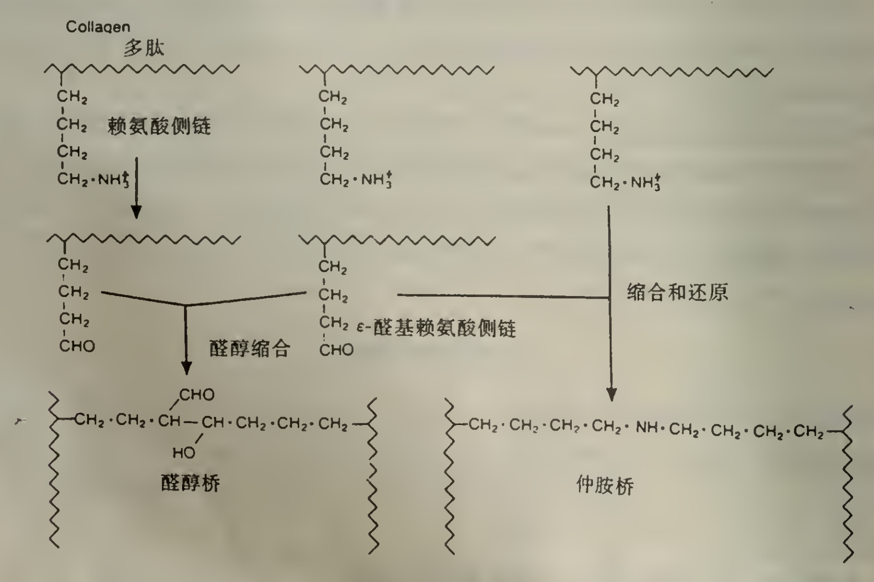


图 9-30　胶原蛋白中多肽链间的共价交联

4. 羧化作用　一些蛋白质的谷氨酸和天冬氨酸可发生羧化作用。例如，血液凝固蛋白凝血酶原(prothrombin)的谷氨酸在翻译后羧化成 γ-羧基谷氨酸，后者可以与 Ca^{2+} 螯合。这是由依赖于维生素 K 的羧化酶催化的。

5. 甲基化　在一些蛋白质中赖氨酸被甲基化。在肌肉蛋白和细胞色素 c 中含有一甲基和二甲基赖氨酸。大多数生物的钙调蛋白含有三甲基赖氨酸。有些蛋白质中的一些谷氨酸的侧链羧基也甲基化。

在蛋白质合成中多种氨基酸均可发生修饰，表 9-8 列出各种氨基酸的修饰。

表 9-8 蛋白质生物合成中氨基酸残基的修饰

氨基酸残基	修饰
丙氨酸	氨基酸末端甲基化
精氨酸	ADP-核糖基化；氨基末端甲基化
天冬酰胺	ADP-核糖基化；糖基化；氨基末端甲基化；β-羟化作用
天冬氨酸	在 GPI-锚锭蛋白中以酰胺连接于乙醇胺；β-羟化作用
半胱氨酸	二硫键形成；脂肪酰化作用
谷氨酸	γ-羟基化作用；甲基化作用
谷氨酰胺	交联赖氨酸的氨基；氨基末端甲基化；内部环化成氨基末端焦谷氨酸
甘氨酸	转变成羧基末端酰胺；氨基末端的肉豆寇酰化
组氨酸	形成白喉酰胺(dipbthamide)，以后 ADP-核糖基化；氨基末端甲基化
赖氨酸	羟化作用后 5-羟赖氨酸糖基化；交联形成；乙酰化作用
甲硫氨酸	氨基末端甲酰基团脱甲酰化作用；氨基末端甲基化
苯丙氨酸	氨基末端甲基化
脯氨酸	羟化作用形成 3-或 4-羟脯氨酸；氨基末端甲基化
丝氨酸	磷酸化作用；糖基化作用；脂肪酰化作用；在 tRNA 水平上硒代半胱氨酸的形成
苏氨酸	磷酸化作用；糖基化作用；脂肪酰化作用
酪氨酸	磷酸化作用；哺乳动物 α-微管蛋白中羧基末端残基的交换

（四）糖基化

在多肽链合成过程中或在合成之后常以共价键与单糖或寡糖侧链连接，生成糖蛋白。这些糖可连接在天冬酰胺的酰胺上（N-连接寡糖）或连接在丝氨酸、苏氨酸或羟赖氨酸的羟基上（*o*-连接寡糖）。糖基化是多种多样的，可以在同一条肽链上的同一位点连接上不同的寡糖，也可以在不同位点上连接上寡糖。糖基化是在酶催化反应下进行的。糖蛋白是一类重要的蛋白，许多膜蛋白和分泌蛋白均是糖蛋白。

（五）脂类对蛋白质的共价修饰

在生物膜中，蛋白质埋藏在脂类双层内。脂类可以非共价键与蛋白质的疏水区相互作用，也可以共价键将脂肪酸连接于蛋白质。蛋白-脂的连接有 3 个主要类型：

(1)在翻译后，脂肪酸与半胱氨酸、丝氨酸或苏氨酸侧链酯化，如髓磷脂的蛋白脂质、病毒糖蛋白、人类铁传递蛋白受体等均是。参与酯化的脂肪酸有肉豆蔻酸、棕榈酸、硬脂酸、油酸等，以脂酰-CoA 为供体。

(2)在翻译中，连接肉豆蔻酸于 N-端甘氨酸。大多数 N-肉豆蔻酰化蛋白质与脂结合，也有少量在细胞溶质。

(3)在翻译后，通过乙醇胺将糖基-磷脂酰基肌醇(GPI)连接于多肽前体的接近 C 端的氨基酸残基上，生成与膜结合的 GPI-锚定蛋白(GPI-anchored protein)。

（六）ADP-核糖基化

在翻译后，一个或多个 ADP-核糖从 NAD^+ 上转移至受体蛋白质上，核糖与氨基酸残基以 N-或 *o*-糖苷键结合。例如，单个 ADP-核糖与精氨酸、天冬酰胺或白喉酰胺(diphthamide，一种修饰的组氨酸)残基的 N 以 N-糖苷键结合。多 ADP-核糖基化主要在细胞核内进行，先将一个 ADP-核糖与谷氨酸残基的羧基侧链或与 C 端赖氨酸的羧基生成 *o*-糖苷连接，以后再连接上多个 ADP-核糖，形成多聚(ADP-核糖)侧链。细胞核内发生多 ADP-核糖基化的蛋白质主要有组蛋白，以及 RNA 聚合酶 α 亚基和拓朴异构酶 I 等。

（七）乙酰化

蛋白质的乙酰化普遍存在于原核生物和真核生物中，例如，烟草花叶病病毒外壳蛋白，在细菌、真菌、动植物中均存在。乙酰化有二类型：一类是由结合于核糖体的乙酰基转移酶催化，将乙酰-CoA 的乙酰基转移至正在合成的多肽链上，当将 N 端的甲硫氨酸除去后，便发生 N-乙酰化，例如卵清蛋白的乙酰化便是如此；另一类型是在翻译后由细胞质的酶催化发生乙酰化，例如肌动蛋白和猫的珠蛋白。此外，细胞核内的组蛋白的内部赖氨酸也可以乙酰化。

（八）磷酸化

酶、受体、介体(mediator)、调节因子等蛋白质的可逆磷酸化是普遍存在的蛋白质修饰，在细胞生长和代谢调节中有重要功能。磷酸化发生在翻译后，由各种蛋白质激酶催化，将磷酸基团连接于丝氨酸、苏氨酸和酪氨酸的羟基上。在磷酸酯酶的作用则发生脱磷酸作用。

（九）C 端酰胺基的引入

一些蛋白质如肽激素含有 C 端酰胺基。这是由多肽链的 C 端甘氨酸的氧化修饰生成的。在肽基甘氨酸羟化酶作用下，将甘氨酸的 α-C 羟化，以后甘氨酸以乙醛酸态除去，并歧化为酰胺。

（十）蛋白质酪氨基的硫酸化

真核生物的蛋白质的酪氨酸普遍发生硫酸化。由酪氨酰蛋白质转移酶从 3′-磷酸腺苷-5′-磷酰硫酸将硫酸基转移至特定的酪氨酸残基上。

第八节　蛋白质的易位与分泌

许多蛋白质要从它们合成的地方转运至其它位置。在真核生物，大多数蛋白质是由细胞质核糖体合成，许多酶包括在代谢中的特定的蛋白质，如血红蛋白与肌肉蛋白是保留在细胞质，而其它蛋白质则易位进入细胞器与膜、或是分泌出细胞。虽然，线粒体和叶绿体能合成少量的蛋白质，仍需要从细胞质转移入蛋白质。原核生物缺乏细胞器，它合成的蛋白质有些特定留在细胞质内，而另一些则要插入到膜，或是分泌出去(图 9-31)。

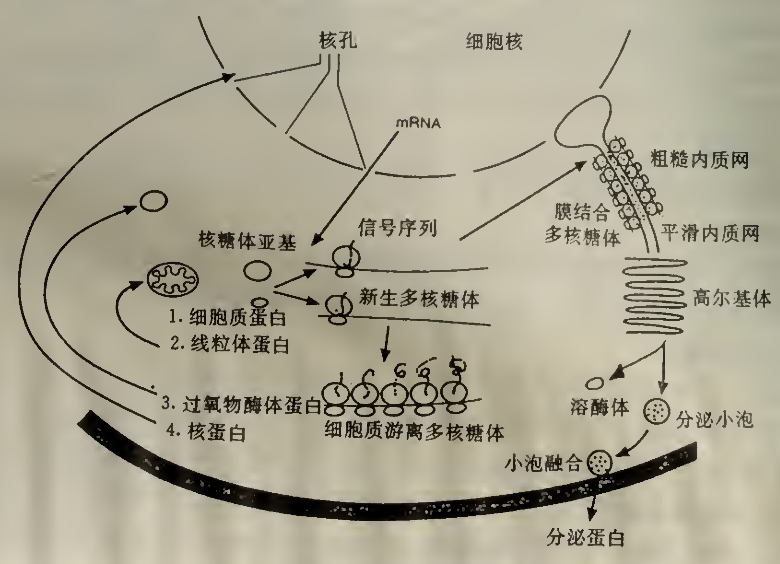


图 9-31　真核生物蛋白质易位的主要途径

在真核生物中，蛋白质从细胞排出以及易位到细胞器与在多肽链或其前体上的特定信号序列（信号肽）有关。这些信号肽指导多肽从细胞内一个区域转运至另一区域，或使之停留在特定的位置。新生的多肽链可以含有一个以上的信号肽，使之易位至某一位置，然后使之停留在该位置。

一、蛋白质易位的途径

蛋白质的易位有 2 个主要途径。

（一）共翻译（co-translation）易位

由与内质网（粗糙内质网）结合的核糖体合成的蛋白质前体，在其 N 端含有一个信号序列（信号肽），它使新生肽链在合成过程中（即共翻译）插入到内质网膜上的特殊通道，然后转移入其内腔。在切去信号肽后，蛋白质可以停留在内质网腔内，或进一步通过高尔基体转移至溶酶体，或分泌小泡，然后分泌出去。蛋白质停留在某一给定位置是受分子上的信号肽控制的，它防止蛋白质进一步的运送。

（二）翻译后易位

是由游离核糖体合成的蛋白质前体在多肽链合成之后，将蛋白质从细胞质转移到线粒体与叶绿体等细胞器中去。

二、蛋白质的共翻译易位

（一）新生肽进入内质网

定位在膜中的、溶酶体的或分泌的蛋白质前体的 N 端含有信号肽。现已有上百种信号肽被测序。信号肽含约 13～36 个氨基酸残基，在或靠近其 N 端有一至多个带正电荷的氨基酸，在信号肽的中部为由 10～15 个氨基酸（大部或全部是疏水性的）组成的疏水核。在信号肽的 C 端靠近断裂位点亦有一短段序列，含侧链较短的和较具极性的氨基酸（如丙氨酸）（图 9-32）。疏水核有助于新合成的肽链附着于内质网的膜上。少数蛋白如卵清蛋白含有和信号肽功能相同的内序列，但在易位后不被切除。

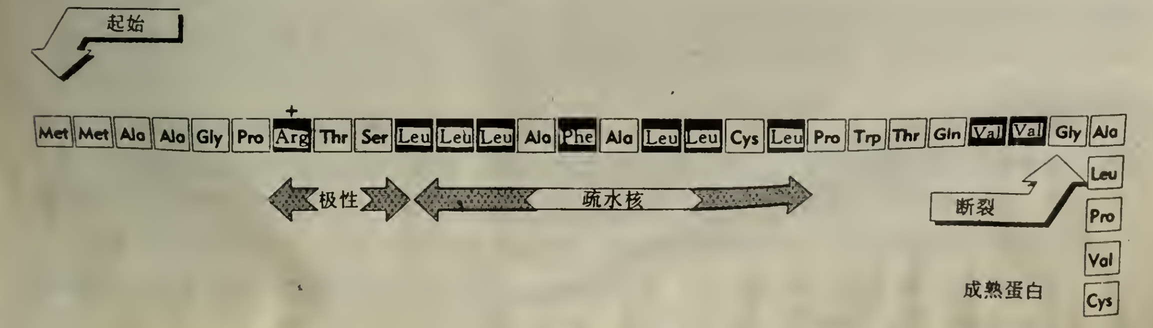


图 9-32　牛生长激素的 N 端由 29 个氨基酸组成的信号序列，含中心疏水核，其前面是极性氨基酸

一旦新生肽链从核糖体出现并延伸时，信号肽便被信号识别颗粒（signal recognition particle，SRP）所识别（图 9-33）。SRP 与携带新生多肽链的核糖体相互作用，引起翻译暂时的终止。SRP 的功能就是将暂停翻译的新生肽链和内质网膜靠近。随后，SRP-核糖体与内质网上一个 SRP 受体（又称停泊蛋白，docking protein，DP）结合，通过一个 GTP 依赖过程，打开一个通

道。另一个内质网膜蛋白即信号序列受体(signal sequence receptor,SSR)与信号肽结合,并促进新生多肽链进入易位通道。另外,在内质网膜上,还有核糖体的受体.这样,通过这3个受体的多重识别,信号肽可以无误地与内质网专一地结合。同时,SRP释放入细胞质,多肽链易位到内质网的内腔。被释放出的SRP则再用于另一个蛋白质的转运。信号肽在多肽链合成完成之前,即由内质网内的信号肽酶切除掉。在易位时会发生多肽链的修饰,如天冬酰胺残基的N-糖基化作用,或是在膜蛋白的脂肪酸的酰基化作用。

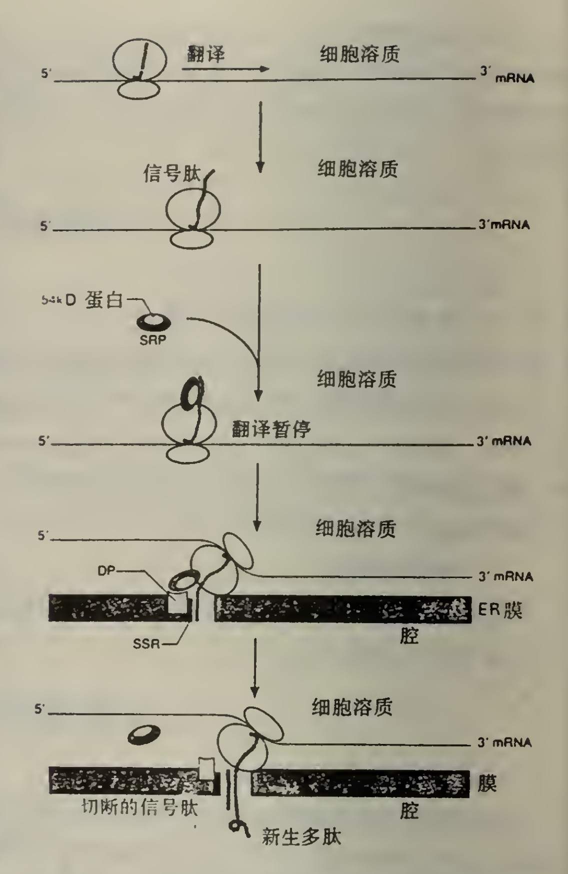


图9-33 蛋白质的共翻译转移入内质网内腔
SRP,信号识别颗粒;DP,停泊蛋白(信号识别颗粒受体);SSR,信号序列受体;ER,内质网。

SRP是由1种含300个核苷酸的TSL-RNA与6个不同多肽链组成。6个不同多肽链的分子量分别是9 kD,14 kD,19 kD,54 kD,68 kD,72 kD。头2个组分的功能是抑制翻译。含有一个GTP结合结构域的54 kD蛋白的功能是与信号肽相互作用。68 kD/72 kD是杂二聚体,由它们与停泊蛋白结合。停泊蛋白含有2个亚基,70 kD的α与30 kD的β多肽链。α亚基整合于内质网膜内,它的一级结构中,有多个疏水区,中间是一个高度正电荷的肽段。正是此肽段与SRP的7SL-RNA作用,使SRP结合于α亚基上。α亚基具有GTP/GDP的结合能力,这一结合能力调节了停泊蛋白与SRP的结合。β亚基的功能不清楚。SSR为一个六聚体,含有各为34 kD与23 kD的糖蛋白。它们可以与信号肽和新生肽链的其它部分相互作用。促进新生肽链穿越内质网,它们还可以在新生肽链易位与切除信号肽过程中起作用。

现知SRP对翻译还有负调节(negative regulation)作用。哺乳动物分泌蛋白质中的许多种都是降解酶类(如核酸酶、蛋白酶等),它们的偶然出现于胞浆内都会造成细胞内的灾难。SRP暂时终止这些蛋白质的翻译,确保这些蛋白质未到达内质网膜之前不会完成翻译。这样,在信号肽和SRP的共同作用下,使得这些分泌蛋白能及时进入内质网腔内,完成转运和分泌。

现知在新生肽链的C端也存在一些影响肽链穿越内质网膜的肽段,使穿越膜的过程终止。这部分称为终止转移序列,也可认为是一种信号序列。

一般地说,新生肽链进入内质网是一个共翻译过程,但也有例外,如在酵母菌内,蛋白质是在翻译后进入内质网的。

(二)蛋白质在内质网内的滞留

进入内质网腔的蛋白质一部分滞留在内质网内,但大多数蛋白质则在内质网腔内被加工,然后转入高尔基体,最终转送到细胞其它位置,或是由胞泌作用被排出。如果某些分泌蛋白不

具有正确的空间构象，肽链是不能通过内质网的，此时，常常会被重链结合蛋白（热休克蛋白的一种）Bip 结合，先被滞留然后进一步会被降解。现在知道，定位在细胞内的蛋白质是与一些特定的肽链序列有关。滞留在内质网的蛋白质，如重链结合蛋白（Bip），2 个葡萄糖调节蛋白（grp 78 与 grp 94）以及二硫键异构酶，在它们的羧基末端找到了四肽 KDEL（即 Lys · Asp · Glu · Leu）序列。如果除去重链结合蛋白 C-末端的 KDEL，它就被分泌。溶菌酶的 C-末端接上了这一四肽，溶菌酶就被保留在内质网中，不再被分泌。在酵母中也有类似的情况，只是四肽的序列是 HDEL（H＝His）。在内质网内的蛋白质有一部分插入到膜内，成为膜的整合蛋白（integral membrane protein）。这些肽链有一段锚定（anchor）序列，作为终止转移信号（stop-transfer signal）。它由一些疏水性氨基酸组成。它的作用是使肽链插在膜上，而不会整条穿越过去。膜上的蛋白有 2 个类型：Ⅰ 型是 N 端在内侧，C 端在外侧（细胞溶质一侧）；Ⅱ 型则相反，C 端在内侧而 N 端在外侧。大部分膜蛋白属于 Ⅰ 型，只有少数属于 Ⅱ 型。

（三）蛋白质在内质网内的加工

在内质网内的蛋白质发生两种重要的变化：折叠和糖基化修饰；修饰与折叠是相联系的，正确的折叠必须加上糖基。同时，在蛋白质二硫键异构酶的作用下，发生二硫键的重排。

重链结合蛋白 Bip 参与了折叠过程。它是属于 Hsp70 家族的一种分子伴侣。它在内质网腔内促进蛋白质寡聚化和/或折叠。蛋白质的寡聚化对以后的转运是重要的。Bip 也起着除去错误折叠的蛋白质的作用。

蛋白质的糖基化是在内质网和高尔基体内进行的。糖基化包括下列步骤：

（1）在内质网的膜上有一种高度疏水性的脂质长萜醇（lipid dolichol），它以焦磷酸基连接 2 个 N-乙酰葡萄糖胺，9 个甘露糖和 3 个葡萄糖残基。

$$\mathrm{HO{-}CH_2\cdot CH_2{-}\underset{\displaystyle H}{\overset{\displaystyle CH_3}{C}}{-}CH_2{-}(CH_2{-}CH{=}\overset{\displaystyle CH_3}{C}{-}CH_2)_n{-}CH_2{-}CH{=}\overset{\displaystyle CH_3}{C}{-}CH_3}$$

长萜醇（$n=9\sim22$）

（2）在内质网膜上的糖基转移酶的催化下，将上述糖基转移至膜上的多肽链的天冬酰胺残基上，形成 N-糖苷键连接的糖蛋白，其糖基暴露于内质网腔。

（3）在内质网内的糖苷酶作用下，依次切除 3 个葡萄糖残基和 1～4 个甘露糖残基。

（4）以后将生成的糖蛋白转移至高尔基体，并在其中由甘露糖苷酶再切除几个甘露糖残基，连接上 N-乙酰葡萄糖胺、半乳糖、唾液酸等，完成糖基化，生成复合糖蛋白：

```
  ⌇
  NH                                    Man—AcG—Gal—SiA
  |              H                     /   \
  HC—CH2—C—N—ACG—ACG—Man          AcG—Gal—SiA
  |           ‖                        \
  C=O         O                         Man—AcG—Gal—SiA
  ⌇
肽链   天冬酰胺    └──────内核──────┘  └──端区──┘
```

AcG：N-乙酰葡萄糖胺；Man：甘露糖；Gal：半乳糖；SiA：唾液酸。

高尔基体是有极性的，其顺式侧朝向内质网，反式侧朝向质膜。上述的糖基化步骤是严格依次在高尔基体内各个囊泡(cisterna)之间由顺式侧向反式侧方向逐步进行的。当到达反式侧之后，便分别转移到不同的终点位置。

(四)蛋白质的定位和分泌

蛋白质进入高尔基体后，便分别运送至不同的目的地，如溶酶体、质膜或分泌出细胞外。蛋白质是通过膜小泡(vesicle)的作用，在其中运送的。蛋白质一旦进入内质网的膜后，便一直保留插入膜内，直至到达最后目的地。

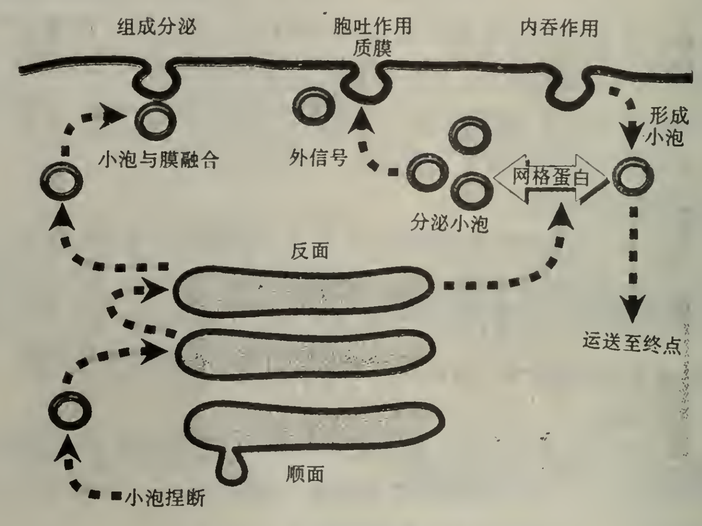


图 9-34　蛋白质在小泡内的运送

小泡是由膜形成的，膜先是凸起，“出芽”，以后捏断，便形成圆形小泡(图 9-34)。小泡的形成要求小分子的 GTP 结合蛋白和消耗 ATP。小泡外面含有一层蛋白外壳(coat)，其内是一层称为接合体(adaptor)的蛋白，内层是包含有待运送的蛋白质膜(图 9-35)。小泡由膜上脱下后，包上外壳，移至靶膜上，脱去外壳，然后与靶膜融合，其中的蛋白质即运送到该靶膜上去。

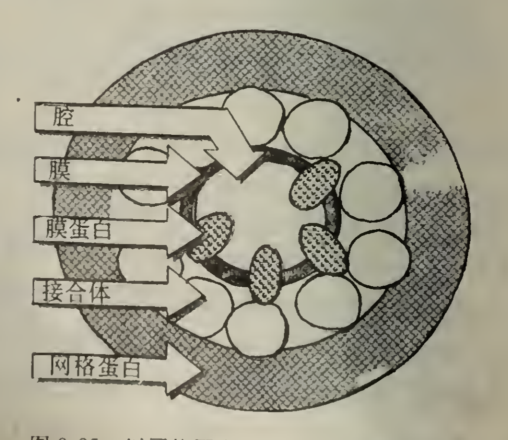


图 9-35　以网格蛋白为外壳的小泡，外壳外层为网格蛋白，内层为接合体蛋白

小泡视其外壳蛋白的不同而分为两个类型，分别负责两个不同途径的运送。一个类型的小泡由一种未知的蛋白做外壳，负责由内质网⟶高尔基体⟶质膜的运送，称为组成分泌(constitutive secretion)或大量流动(bulk flow)；另一类型的外壳蛋白是网格蛋白(clathrin)，称为分泌小泡(secretory vesicles)，由高尔基体的反式侧将蛋白质通过胞吐作用(exocytosis)分泌出去。细胞外的蛋白质也可以由这种小泡通过内吞作用(endocytosis)运入细胞内(图 9-34)。

溶酶体是由膜包着的小体，内含多种水解酶。蛋白质可以由高尔基体反式侧生成的小泡，

再经过内体(endosome 是一种小管和小泡组成的外包以膜的结构),进入溶酶体。在蛋白质糖基化过程中形成的甘露糖-6-磷酸是作为蛋白质进入溶酶体的识别信号。细胞外的蛋白质也可以通过内吞作用,由小泡运送,通过内体而进入溶酶体。

三、蛋白质的翻译后易位(translocation for post-translation)

细胞器例如线粒体、叶绿体、细胞核、过氧物酶体的许多组成蛋白质是由游离的核糖体合成的,并作为前体释放到细胞质中去,随后为细胞器所接受,最终成为结构蛋白质。由核糖体合成的蛋白质前体按其所携带的信号不同而分别转运到不同的细胞器中去(表 9-9)。可看出,导向不同细胞器的信号序列的位置、性质和长短都不同,如过氧物酶体蛋白的信号只有 3 个氨基酸残基。如果缺乏这些信号,合成的蛋白质便停留在细胞溶质中,呈似溶解(quasi-soluble)态。

表 9-9　将游离核糖体合成的蛋白质导向不同细胞器的信号

细胞器	信号位置	类型	信号长度(残基数)
线粒体	N-末端	带电荷	12～30
叶绿体	N-末端	带电荷	～25
细胞核	内部	碱性	7～9
过氧物酶体	C-末端	SKL*	3

* SKL:Ser-Lys-Leu

(一)蛋白质进入线粒体

对酵母线粒体的研究表明蛋白质的进入有多条途径(图 9-36),其中的途径 A 可分为 4 个阶段:

(1)前体蛋白结合于特定的进口位点,它是由 42 kD 的蛋白质组成,位于线粒体的外膜。在这个区域,外膜与内膜是靠近的;

(2)蛋白质与一种抗折叠蛋白质相结合,使之解折叠;

(3)易位穿过内膜,这个过程需要 ATP 的水解和在膜外侧要有一个正电荷的电化学梯度;

(4)受溶解于衬质的金属蛋白酶的作用,切除 N 端的序列,并折叠成三级结构,分子伴侣和 ATP 起促进作用。

途径 B 包括了蛋白质进入线粒体外膜,是在细胞质中完成的多肽链合成之后直接插入。

定位于内外膜之间的蛋白质,可以有两条途径由细胞质进入。一条途径是 C,蛋白质插入到外膜脂双层,在内外膜空间进行分子修饰。例如,细胞色素 c 作为成熟的脱辅基蛋白质穿过外膜;然后,在存在于外膜内侧的细胞色素血红素裂合酶的催化下,将它在膜间的空间处转变为与血红素共价结合。血红素蛋白质随后插入到内膜的外侧。另一条途径是 D,例如,细胞色素 b2 开始是将前体插入衬质(途径 A),然后分两步切除前序列,先是在衬质,其次是内膜。成熟的蛋白质就释放到两层膜中间去。

蛋白质进入线粒体内不同部位(内、外膜、内外膜间隙、衬质)是由蛋白质前体的 N 端前序列(信号序列)决定的。前序列的前一半使蛋白质穿过两层膜进入衬质,称为“衬质靶向信号”(matrix-targeting signal)。前序列的第二部分称为“膜靶向信号”(membrane-targeting signal)。在蛋白质进入衬质后,由在其中的 Mg^{2+} 依赖蛋白酶切去衬质靶向信号。如果前序列中没有其它信号,蛋白质便停留在衬质中;如果后一半有膜靶向信号,在切去前一半后,膜靶向

信号(此时为N端)便使蛋白质转送到外膜、内外膜间隙、或内膜。最后再切去此信号序列。此外,也可以由一个“终止转移序列”(stop-transfer sequence)使蛋白质停留在外膜或内膜上。

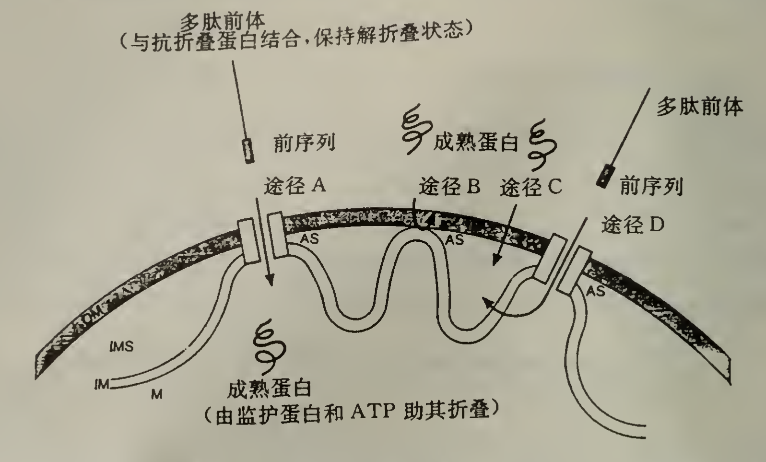


图 9-36　蛋白质易位入线粒体。蛋白质从细胞质通过不同途径进入线粒体的不同区域
AS 线粒体内膜和外膜的附着点;OM 外膜;IMS 膜内空间;IM 内膜;M 衬质。

蛋白质前体首先在线粒体外膜上与受体蛋白结合。已知在酵母线粒体外膜上有3种受体蛋白,分别称为$Tom^{20}p$,$Tom^{22}p$,$Tom^{70}p$。以后蛋白质前体即通过跨膜的通道进入线粒体。这通道是由膜的整合蛋白组成的,具有亲水性。一种称为“蛋白质易位马达”(protein translocation motor)由水解核苷三磷酸驱动蛋白质通过通道,跨膜电位也有助于蛋白质通过。

(二)蛋白质进入叶绿体

蛋白质进入叶绿体的过程和进入线粒体相似,其前序列由约50个氨基酸组成,它的前一半指导蛋白质进入叶绿体间质,然后在第20~25个氨基酸处切断,余下后半段即指导蛋白质进入类囊体膜或腔。蛋白质前体结合于叶绿体表面受GTP控制。有两种具有结合GTP的基序(motif)的蛋白质Oep34和Oep86认为可能是蛋白质进入的受体。

(三)蛋白质进入细胞核

细胞核内全部的蛋白质都是由细胞质输入的。在细胞核的核膜上有核孔,是细胞核与细胞质交换大分子的通道。高等真核生物的核孔复合物由约100种不同多肽组成,其质量达125×10^3 kD。但其孔径只有约9 nm,分子较小的蛋白质如细胞色素c可以自由扩散通过,较大分子的牛血清清蛋白则不能通过,所以,大分子的蛋白质进入细胞核是一个主动过程,而且要求有信号指引。细胞核定位序列(nuclear localization sequence,NLS)由一至多个碱性氨基酸簇组成。蛋白质进入细胞核分二步:先是携带NLS信号的蛋白质与在细胞质中的“输入蛋白”(importin,又称karyopherin或NLS受体)结合。输入蛋白由α,β2个亚基组成。蛋白质先与α亚基结合,然后通过输入蛋白-β结合域(importin-β binding domain, IBB)与β亚基结合。由β亚基介导此蛋白复合物停泊于核孔复合物。下一步是在小GTP酶Ran(small guanosine triphasphatase, Ran)作用下水解GTP,并使蛋白复合物通过核孔进入核内,输入蛋白的2个亚基则释放出再回到细胞质。整个过程如图9-37所示。

(四)蛋白质进入过氧物酶体

大多数输入过氧物酶体的蛋白质均具有C端的Ser-Lys-Leu序列或其十分近似的序列,

统称为过氧物酶体靶向信号1(peroxisome-targeting signal 1,PTS1),另一些少数蛋白则具有N端的信号,是含11个残基的共有基序(consensus motif),称为PTS2。更有其它蛋白质含未鉴定的信号。PTS1信号被称为Pas8p或Pas10p的蛋白识别,视酵母种的不同而异。PTS2信号被Pas7p蛋白识别。上述的Pas8p,Pas10p和Pas7p是在细胞质中还是在膜上仍有争论。关于蛋白质进入过氧物酶体的机制及其动力仍不清楚。

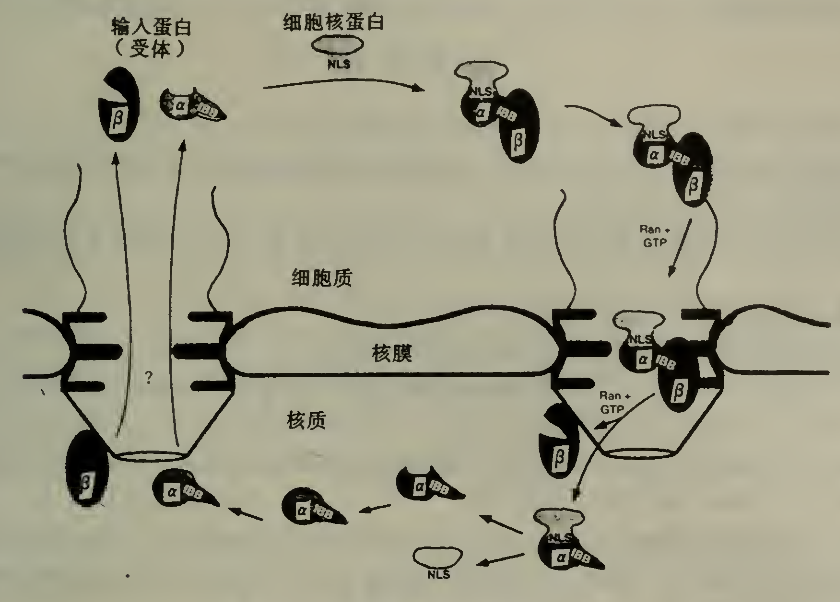


图 9-37 蛋白质输入细胞核的步骤

四、细菌的蛋白质易位

上文所述的是真核细胞的蛋白质易位,在原核细胞(细菌)内也有类似情况,蛋白质在细胞质合成后要转移入内膜和进入周质空间,有时还通过外膜。在大肠杆菌内,一般是共翻译易位,但也有少数蛋白质是可以同时共翻译和翻译后易位的。

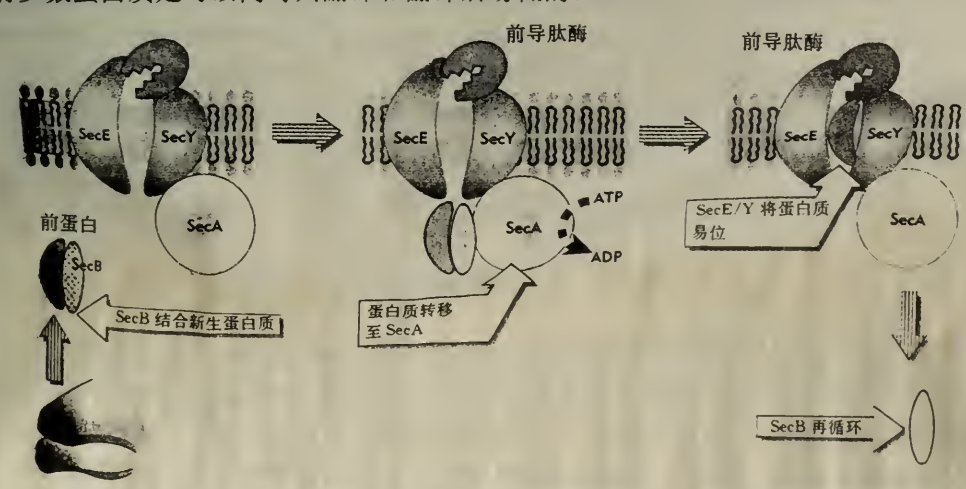


图 9-38 细菌内蛋白质的易位

细菌分泌的蛋白质含有N端的前导序列，其前端为亲水性，连接一疏水核。蛋白质进入膜后的正确定位也要求次级信号。蛋白质的正确构象是关键。已知有多种分子伴侣参与蛋白质的折叠，包括SecB，SecA，SecE，SecY等。新生肽链先与SecB结合，使之不会发生错误折叠，并把肽链导向膜上的SecA。SecA是一个大的膜周边蛋白，它同时具有ATP酶活性。SecA与蛋白质和ATP结合，并水解ATP。以后肽链即由SecA转移给SecE/Y（均是膜的整合蛋白）。最后由膜上的前导肽酶(leader peptidase)切去蛋白质的前导序列。上述过程如图9-38所示。

参考文献

1. 徐铃，彭朝晖. 基因的分子生物学. 北京：中国科学技术出版社，1992. p. 107～154.
2. 凌俊，刘望夷. 核糖体的拓扑结构及蛋白质生物合成（敖世洲主编. 基因分子生物学研究进展）. 上海：上海科学技术出版社，1992. p. 109～116.
3. 冯佑民，张友尚. 蛋白质折叠（敖世洲主编. 基因分子生物学进展）. 上海：上海科学技术出版社，p. 125～132.
4. 邹承鲁. 新生肽链的折叠. 生命科学. 1993. 4：1～6.
5. 王克夷. 肽链翻译后的加工. 生命的化学. 1995. 15(3)：1～4.
6. Arnstein, H. R. V., Cox, R. A. Protein Biosynthesis. New York: Oxford University Press, 1992. p. 19～103.
7. Lehniger, A. L., Nelson, D. L., Cox, M. M. Principles of Biochemistry. (2nd ed.). Chapters 26. New York: Worth Publishers, Inc. 1993. p. 893～938.
8. Stark, H., Mueller, F., Orlova, E. V., Schatz, M., Dube, P., Erdemir, T., Zemlin, F., Brimacombe, R., Van Heel, M. The 70 S *Escherichia coli* Ribosome at 2.3 nm Resolution: Fitting the Ribosomal RNA. Structure. 1995. 3：815～821.
9. Schatg, G., Dobberstein, B. Common Principles of Protein Translocation Across Membranes. Science. 1996. 271：1 519～1 526.
10. Görlich, D., Mattaj, I. W. Nucleocytoplasmic Transport. Science. 1996. 271：1 513～1 518.
11. Lewin, B. Genes V. Part 2. Translation: Expressing Genes as Protein. New York: Oxford University Press, 1994. p. 161～276.
12. Watson, J. D., Hopkins, N. H. . Roberts, J. W., Steitz, J. A., Weiner, A. M. Molecular Biology of the Gene. (4th ed.). Chapters 13. California: The Benjamin/Cummings Pub., Co. Inc. 1987. p. 360～381.

（杨礼锐　李明启）

第十章　原核生物基因表达的调控

第一节　引　　言

基因表达(gene expression)就是指某一基因指导下的蛋白质合成，蛋白质是基因表达的产物。一种生物含有大量基因，例如人体约含有 10^5 个基因。在生活中并非所有基因都一齐表达，而是有些基因进行表达，形成其基因表达的特异产物，以构成细胞结构或代谢所需要的蛋白质或酶类。但是，有许多基因却被关闭，不进行表达，而要在适当的时候才进行表达。

在生物代谢过程中所需要的各种酶或蛋白质的基因以及构成细胞结构成分的基因，在正常情况下是经常进行表达的，而与生物发育过程有关的基因则要在特定的时间或空间才进行表达，而在其余的时间或空间这些基因则被关闭。例如，昆虫变态过程的各种基因要在一定发育阶段(如幼虫、蛹、成虫)才能表达；与植物开花有关的基因也只有在开花前才能表达，而在营养体(根、茎、叶)生长时则全被关闭。

虽然一种基因编码一种蛋白质，但是不同蛋白质在细胞中的相对数量差别很大，随着它们的功能而不同，例如，在 *E. coli* 细胞中，从总蛋白的不足 0.01%～2%，各种蛋白质变化不定。

细胞要使其蛋白质合成达到这种差异，可以有两条途径。第一条途径是细胞控制从其 DNA 模板上转录其特异的 mRNA 的速度，这是一种最经济的办法，可以免去浪费从 mRNA 合成蛋白质的各种元件和材料。这大概是生物在长期进化过程中自然选择的结果。这种控制通常称之为转录水平(transcriptional level)的调控。大多数基因表达都属于转录水平的调控。第二条途径是在 mRNA 合成后，控制从 mRNA 翻译成多肽链的速度，包含一些分子装置问题，如与核糖体的结合速度等。这种蛋白质合成或基因表达的控制称为翻译水平(translational level)的调控。这种调控是较少的。

对于细胞来说，它不能合成超过它所需要的 mRNA，对它才不浪费原料和能量。因此，我们也应当更多地注意转录水平的调控问题。原核生物特别是大肠杆菌(*E. coli*)基因表达调控的原理研究得比较成熟和深入，本章我们以原核生物为主，讨论有关它们的基因表达的调控问题。至于真核生物基因表达的调控则要在第十三章中详细讨论。

第二节　细菌对营养的适应

除光合自养细菌外，一般细菌如大肠杆菌所需的碳源首先是葡萄糖，利用葡萄糖发酵获得能量，维持生存。在缺乏葡萄糖时，细菌也可以利用其它糖类如乳糖作为碳源，维持生活。从大肠杆菌染色体的长度可以估计它大约可以编码 2 000～4 000 种蛋白质。一般估计生长在以葡萄糖为唯一碳源的细菌中的酶类至少有 600～800 种。在这些酶中特别是与降解葡萄糖第一步有关的酶(即所谓关键酶)，及与制造氨基酸与核苷酸有关的酶，以相当大的数量存在着，产生

ATP 所需的酶类也大量存在。与此相反，其它制造少量辅酶的酶类则只以微量存在。用于构建细胞壁、细胞膜及核糖体的各种结构蛋白也必须大量存在。

了解在大肠杆菌细胞中稀有的及丰富的蛋白质有多少个拷贝(copy)是很有必要的。β-半乳糖苷酶(β-galactosidase)是研究得最透彻的酶，它能将乳糖分解成葡萄糖及半乳糖。β-半乳糖苷酶的分子量为 460 kD，具有四聚体结构，由 4 个分子量各为 116 kD 相同的多肽链组成。这是一个极重要的酶，因为乳糖不能用做碳源或能源，它必须先被分解成简单的葡萄糖和半乳糖。生长在以乳糖为唯一碳源的大肠杆菌中通常含有 3 000 个 β-半乳糖苷酶分子，占其总蛋白质的 1.5%，这是所能合成的最大数量。

已经精确计算过，其它丰富的蛋白质包括有蛋白质合成过程的组分——核糖体蛋白(平均分子量 20 kD)等共计 53 种，它们总起来占细胞总蛋白的 20%，每种蛋白质约有 2 000 个拷贝，大肠杆菌中最丰富的蛋白质是蛋白质合成的延长因子 EF-Tu(见第九章)，每个细胞中存在 100 000 个分子左右。

与这些永久分子不同，调节蛋白及酶在每个细胞中只以少数 100～200 个拷贝存在，因此，它们很难检测和纯化，目前我们尚缺乏其数量的确切资料。

当细胞需要一种蛋白质时，与它不需要的时候比较，它能以极大数量存在。例如，*E. coli* 生长在乳糖中，则 β-半乳糖苷酶分子几乎有 3 000 个之多，而生长在其它碳源的细胞此酶只有 1/1 000。像以乳糖为底物那样，引入培养基中而专一地增加某种酶的数量，这个底物即称为诱导物(inducer)，而这种酶则称为诱导酶(inducible enzyme)。

在生物合成中所包含的酶类却有与此相反的反应。例如，生长在不含任何氨基酸的培养基中的大肠杆菌，细胞中含有合成 20 种氨基酸所需要的全部酶类。而当培养基中含有某种氨基酸时，则该种氨基酸生物合成的酶类几乎完全缺失。生物合成的酶类受其终产物(end product)的抑制，如组氨酸为组氨酸生物合成酶类的终产物，此种酶称为阻遏酶(repressive enzyme)。而那些导致某种阻遏酶数量降低的终产物则称为辅阻遏物(corepressor)。诱导与阻遏效应对细菌来说是同等有用的。当需要某种酶转化一种特异的食物分子或者合成一种所需的细胞组分时，它们即存在，而不需要它们时，它们则不存在。

蛋白质含量的变化反映了 mRNA 分子数目的多少。在旺盛分裂的细菌中，大多数蛋白质分子一旦合成后即十分稳定。因此，蛋白质数量的变化通常反映蛋白质合成的速度，而非其相对稳定性，但是也有例外情况。当细胞适应了新的生长条件，某种蛋白质数量的变化常常是 mRNA 模板数量变化的结果。我们估计当 β-半乳糖苷酶合成最多时，每个细胞中存在 50 个 β-半乳糖苷酶 mRNA 分子；而相反在乳糖不存在时，细胞中则只含有不到 1 个 β-半乳糖苷酶的 mRNA 分子。

第三节　组成蛋白与调节蛋白

一、组 成 蛋 白

在细胞内有许多种蛋白质其数量几乎不受外界环境的影响。这些蛋白质称为组成蛋白质(constitutive proteins)。例如，在 *E. coli* 中控制葡萄糖降解的酶类数量并无剧烈变化，不论是将葡萄糖加到培养基中或从培养基中取出，都无大的变化。这可能意味着 *E. coli* 在自然界中

经常得到葡萄糖，以致没有必要演化出一个阻遏产生这些酶类的手段。而且对细菌来说，控制蛋白质极小量的合成也很不经济，如将乳糖阻遏物极低的拷贝数做成组成蛋白质。组成蛋白质合成的速度是遗传上规定的，由于(1)启动子的天然效率，靠它可以制造 mRNA；(2)核糖体阅读信使的速度；(3)其信使被核酸酶降解的敏感性。

二、调 节 蛋 白

调节蛋白(regulatory protein)是一些特殊蛋白质，它们决定着何时诱导酶或阻遏酶可以合成。每种调节蛋白影响一种或多种特殊基因的表达。它们有两种基本类型：即正调节蛋白(positive regulator)及负调节蛋白(negative regulator)。这两种调节可以由调节它们的基因之相反效应加以区别。负调节蛋白或称阻遏蛋白，会使其靶蛋白的合成受到抑制，即不表达，而不管是否需要。这类突变种称为组成突变种(constitutive mutant)。与此相反，正调节蛋白是激活蛋白(activator)，它的存在对合成所调节的酶是需要的，因此，激活蛋白存在时被控制的基因即表达。

第四节　操纵子学说

一、操　纵　子

细菌基因表达调控的许多原理是在研究 *E. coli* 乳糖代谢调节时被发现的。法国巴斯德研究院的 Francois Jacob 与 Jacques Monod 于 1960 年在法国科学院院报(Proceeding of the French Academy of Sciences)上发表了一篇论文，提出乳糖代谢中的两个基因被一靠近它们的遗传因子所调节。这二个基因为 β-半乳糖苷酶(β-galactosidase)和半乳糖苷透过酶(galactoside permease)。前者能水解乳糖成为半乳糖和葡萄糖，后者将乳糖运输到细胞之中。在此文中他们首先提出了操纵子(operon)和操纵基因(operator)的概念，他们的操纵子学说(theory of operon)使我们得以从分子水平认识基因表达的调控，是一个划时代的突破，因此他们二人于 1965 年荣获诺贝尔生理学奖。

Jacob 与 Monod 所提出的关于基因表达调控的操纵子学说可以简述如下：有一个专一的阻遏分子(蛋白质)结合在靠近 β-半乳糖苷酶基因上面，这段DNA 他们称之为操纵基因。由于阻遏分子结合在DNA 的操纵基因上，从而阻止了 RNA 聚合酶合成 β-半乳糖苷酶的 mRNA。此外，他们还指出乳糖为诱导物，当乳糖结合到阻遏分子上时，即阻止阻遏分子与操纵基因的结合。当有乳糖时，阻遏分子即失活，mRNA 就可以转录出来。如果去掉乳糖时，阻遏分子又恢复其活力，与操纵基因 DNA 结合，将乳糖基因关闭。我们可以从下面的图 10-1 了解。在图中上边为乳糖操纵子被阻遏的状态，下边为被诱导的状态。

二、乳糖操纵子

乳糖操纵子(lac operon)是原核生物中研究得最清楚的一种操纵子，它的调节模式如图 10-1。在乳糖操纵子上，除去 β-半乳糖苷酶(*Z*)和半乳糖苷透过酶(*Y*)基因之外，还有一个硫半乳糖苷转乙酰酶(thiogalactoside transacytylase)基因(*A*)(它的生理功能尚不清楚)。这 3 个基

因每个前面都有一翻译信号，引导核糖体结合及蛋白质合成。在底物乳糖不存在时，lac 操纵子基因即被阻遏，β-半乳糖苷酶只以很少的拷贝存在(每个细胞几个分子)。Jacob 与 Monod 发现，操纵基因或一称为 i(或 I)的基因的突变会引起 lac 操纵子基因产物的合成。当 i 基因失效时，将一有功能的 i 基因引入细胞的另一 DNA 分子上阻遏作用可能恢复。这表明：基因编码一个可扩散的分子，引起基因的阻遏。以后这种分子被证明为蛋白质，现在称之为阻遏蛋白(repressor)即上文所说的阻遏分子。阻遏作用并不是绝对的，即使在阻遏情况下，每个细胞也有少数拷贝的 β-半乳糖苷酶及半乳糖苷透过酶。

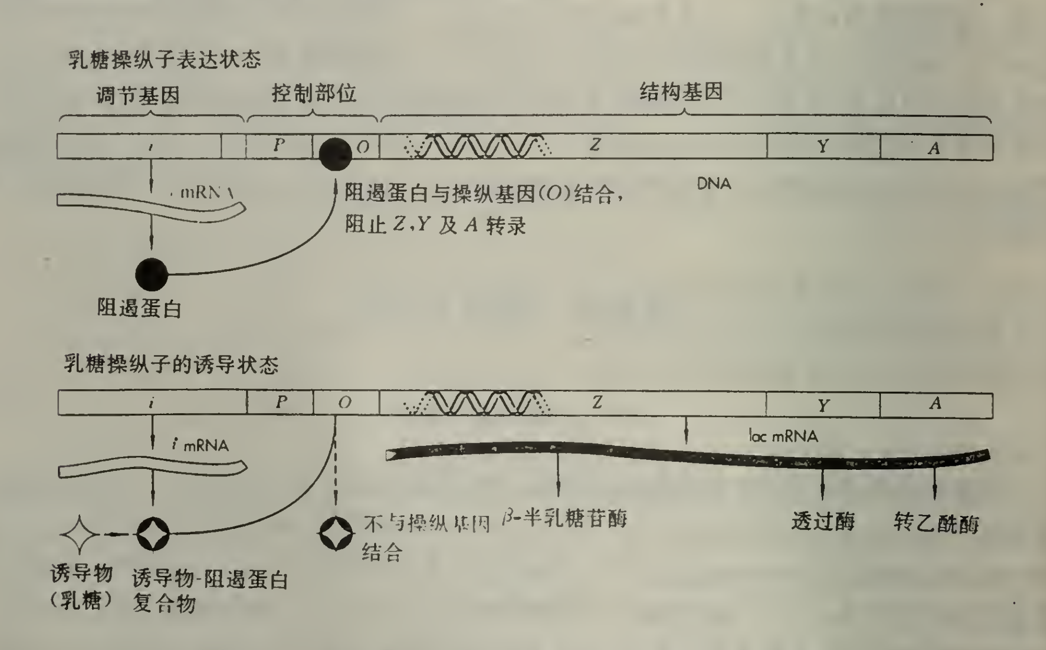


图 10-1　控制乳糖(lac)操纵子有关基因的阻遏蛋白及诱导物

调节基因(i)编码乳糖阻遏蛋白。DNA 链的 P 段为启动子。

当供给细胞以乳糖时，lac 操纵子即被诱导，一个诱导物分子结合在阻遏蛋白的特异部位上，引起阻遏蛋白构象的改变，结果使阻遏蛋白从操纵基因上解离下来(图 10-1)。在此系统中的诱导物并非乳糖本身，而是乳糖的同素异形体，称为异乳糖(allolactose)。乳糖进入 *E. coli* 细胞，被转化成异乳糖。异乳糖与阻遏蛋白结合后，它们即从操纵基因上解离下来，lac 操纵子基因即开始表达，β-半乳糖苷酶的浓度可以增加 1 000 倍之多。

有些结构上与异乳糖相似的 β-半乳糖苷也是 β-半乳糖苷酶的诱导物，但并非它的底物。而某些 β-半乳糖苷为底物，却非诱导物。异丙基硫代半乳糖苷 (isopropyl thiogalactoside，简称 IPTG) 是一种特别有效的诱导物，目前在基因表达的研究中常常使用，在基因表达的系统中加入 IPTG 之后，基因的表达量常成倍地增加。IPTG 不是乳糖代谢的底物，属于非代谢诱导物。

在所有 *E. coli* 的启动子中转录起始部位之前的第 35 及第 10 核苷酸(画框处)都是高度保守的，这两个区域的突变都对转录有明显的影响，突变成单核苷酸或成对的改变，也可能是缺失(图 10-2)。

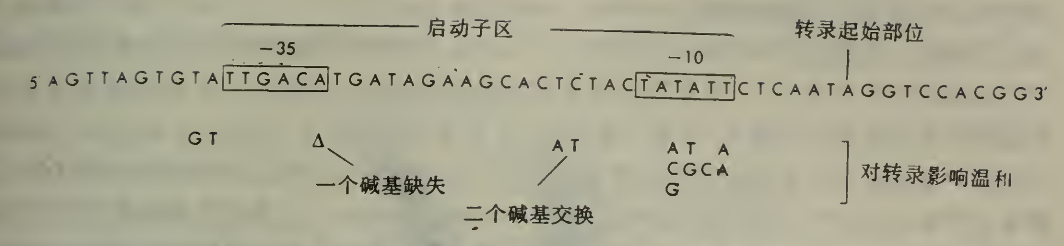


图 10-2　DNA 的特殊序列对 *E. coli* RNA 聚合酶的影响

第五节　操纵基因的结构

一、操纵基因的鉴定

DNA 上蛋白质的结合部位，如一种操纵基因，如何能够鉴定出来呢？原来操纵基因是靠阻

图 10-3　用足迹法鉴定蛋白质在 DNA 上的结合部位

(a)示足迹法如何揭示蛋白质在 DNA 上的结合部位；(b)乳糖阻遏蛋白与操纵基因结合的足迹照片。

(1)乳糖阻遏蛋白与操纵基因结合；(2)DNA 分子被 DNA 酶随机切割后用聚丙烯酰胺凝胶电泳分离；(3) DNA 分子处理如左方，但是用阻遏蛋白与 DNA 结合；(4)核酸酶未分解处表明在核苷酸 110～130 间有阻遏蛋白结合。

遏物的专一结合以免除核酸酶(nuclease)的攻击的,通过DNA片段的分离与序列测定,即可确定其结构。但是这是一件极吃力的工作。由于我们利用限制性内切酶可以确切分离出蛋白质所接触的部位的片段,于是就可以鉴定操纵基因的结构,与测定DNA序列的化学法相似,测定操纵基因的基本原则是:如果一条DNA片段被放射性原子所标记,这条链中任何断裂部位从所产生的标记片段的大小即可推导出来。DNA片段的大小可以用高分辨的聚丙烯酰胺凝胶电泳所确定。这种方法称为足迹图法(foot-printing method),这是仿照用纸层析鉴定蛋白质的指纹图法(finger printing method)命名的。结合部位靠结合蛋白使核酸酶的酶切作用被保护而不被切割(图10-3)。靠这种方法我们可以了解蛋白质在何处与DNA的糖磷酸酯骨架上的碱基及磷酸进行特殊接触。

二、操纵基因的结构

目前已有许多种操纵基因用足迹法定位并进行了DNA序列分析,将影响阻遏物结合的操纵基因突变的特性加以描述,它们最重要的特征是反向重复(inverted repeat)或毗邻重复(near repeat)。例如,用核酸酶消化DNA与lac阻遏物的复合体,分离得到一个乳糖操纵基因片段,由24 bp组成,其中约有16个是与二重对称轴有关,以第11位碱基为中心,其结构如图10-3及图10-4。操纵基因的序列测定出来后就使我们了解到这种对称能够使阻遏蛋白多聚体的两亚基可以同时结合在操纵基因上,从而阻碍了转录的发生。

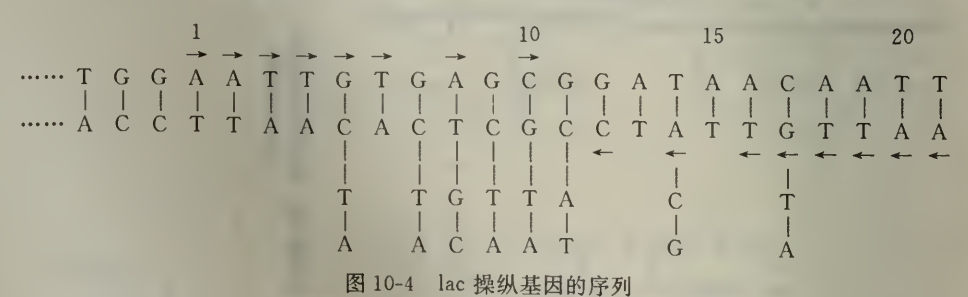


图10-4 lac操纵基因的序列

lac mRNA转录的起点为核苷酸1,核苷酸11为对称的中心,水平的箭头表示操纵基因,另一半具有对称结构。碱基对改变则使lac操纵基因的结合减弱,序列下列出的为野生型。

三、阻遏蛋白与操纵基因的结合

近年来以X射线晶体学测定几种结合蛋白与DNA的三维结构,得出一些惊人的结果。它们表明调节蛋白(阻遏蛋白)与特异的DNA部位是如何结合的。第一个测定的结构是*E. coli* CAP蛋白的结构,随后不久,λCro蛋白、λ阻遏物及噬菌体434的阻遏蛋白(λ的近亲)先后测定成功。有如晶体学家预测的那样,活性结合单位为两个相同的球状多肽链的二聚体,在空间以相反的取向形成一个二重对称轴的构象,与DNA的结合部位的结构相吻合(图10-5)。

调节蛋白与DNA的结合是由两股较短而邻近的α-螺旋形成。λ阻遏蛋白结构是在第2及第3个α-螺旋上,Cro和CAP也有类似结构。在λ阻遏蛋白中螺旋3从蛋白质表面突出,与其对手(即λ阻遏蛋白二聚体第二个亚基)恰好以B-DNA(3.4 nm)的螺旋重复距离相隔开。模型表明螺旋3与大沟恰相吻合,于是阻遏蛋白二聚体通过每个单体的螺旋3与操纵基因的一半相互接触而与DNA的一侧相结合。大部分接触是由大沟的碱基边沿与螺旋3侧链的氢键靠

互补完成。下列事实有力地证实这个模型，即影响λ阻遏蛋白与操纵基因的大多数突变种都是改变两个相邻α-螺旋的氨基酸。最近这个模型已用噬菌体434 X射线分析直接证实。

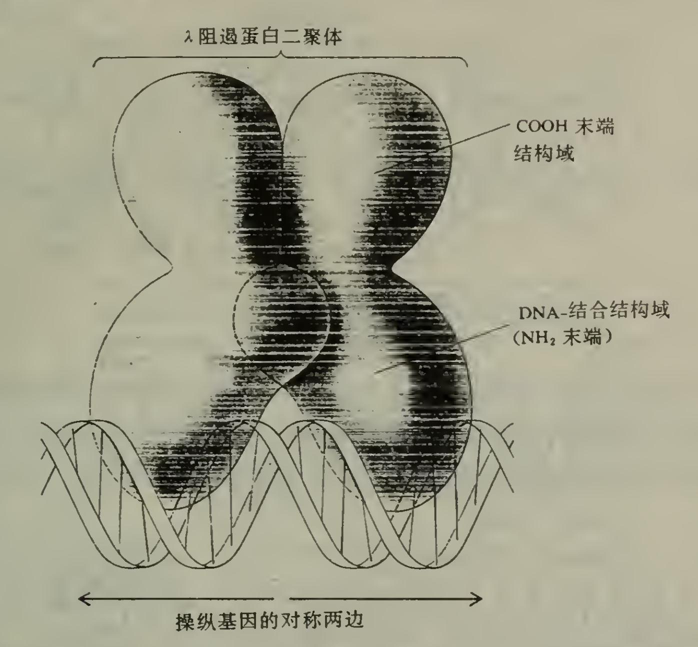


图 10-5　λ阻遏蛋白与操纵基因结合的示意图，示N-末端DNA结合的结构域及C-末端结构域

从这些例子中α-螺旋的结构以及氨基酸序列的知识使我们可以靠检测氨基酸序列就能够预测这个结构是否可以形成。由10种以上DNA结合蛋白的研究结果指出：成对的α-螺旋是结合专一性DNA调节蛋白部位的一般特征。与DNA相结合的调节蛋白的α-螺旋-转折-α-螺旋(helix-turn-helix)的超二级结构(motif)的发现是阐明遗传调节基础的一个重要进展。

螺旋-转角-螺旋是许多原核生物调节蛋白中蛋白质与DNA相互作用的基础。螺旋-转角-螺旋由两个短α-螺旋片段组成，各约7～9个氨基酸残基长，中间由一条β-转角（约20个氨基酸长）分隔开。通常这个结构本身并不稳定，但它代表蛋白质与DNA结合的反应部分。这两个α-螺旋中有一个称为识别螺旋（recognition helix），因为它含有许多能与DNA相互作用的氨基酸。这个螺旋定位于DNA的大沟中，噬菌体434的Cro阻遏蛋白就是一个典型例子（图10-6）。

除去螺旋-转角-螺旋之外，近年发现锌指(Zn finger)和亮氨酸拉链(leucine zipper)也是调节蛋白的DNA结合结构域中起重要作用的超二级结构。

锌指由大约30个氨基酸残基组成，其中的4个残基或者是半胱氨酸(Cys)，或者是两个半胱氨酸(Cys)和两个组氨酸(His)，与一个锌(Zn^{2+})原子配位。这种超二级结构存在于许多原核生物的DNA结合蛋白中。噬菌体T4有一种基因32蛋白(gene 32 protein)，能与单链DNA结合，就具有锌指的结构。

亮氨酸拉链是一种特殊的α-螺旋，它的疏水氨基酸集中排列在螺旋的一侧。疏水的表面是2个蛋白质分子构成二聚体的接触点。这些α-螺旋的显著特点是亮氨酸的频繁出现，它们在每7个氨基酸残基出现1个，使之沿着α-螺旋的疏水侧排列成直线，有如日常使用的拉链结构，故称之为亮氨酸拉链。

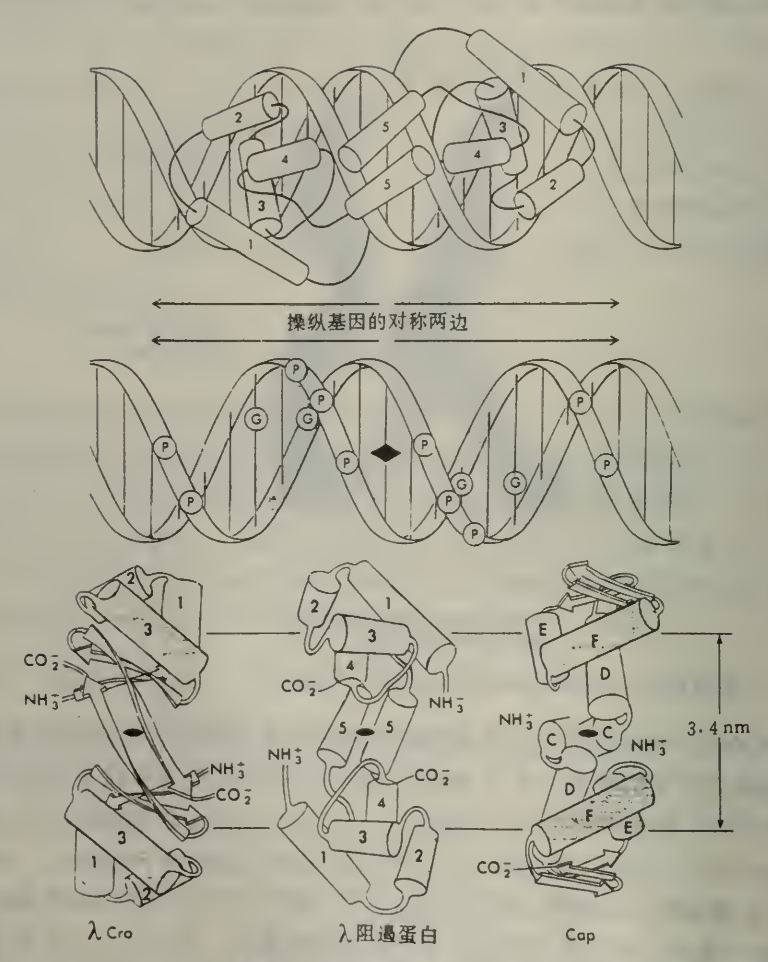


图 10-6　上图：两个 λ 阻遏蛋白单体与操纵基因结合，二者互相适合；下图：三种调节蛋白（λ Cro，λ 阻遏蛋白，Cap）具有相似结构，α-螺旋由 β-转折连接圆柱代表 α-螺旋，两个 α-阻遏蛋白单体在操纵基因中心两边呈对称状结合。

第六节　乳糖操纵子

一、乳糖操纵子的正调控

起初我们认为乳糖操纵子仅仅处于乳糖阻遏物的负控制之下，但是现在了解到，即使有诱导物存在将阻遏蛋白中和，操纵子的正常表达还必须有一个蛋白质介导的正调控信号（图 10-7）。在研究葡萄糖怎样阻断许多种操纵子（葡萄糖敏感操纵子 glucose-sensitive operons）的功能时，发现了这种正调控蛋白的存在，每一种正调控蛋白控制一种糖（如乳糖、半乳糖、阿拉伯糖及麦芽糖）的降解。例如，*E. coli* 生长在既有葡萄糖又有乳糖存在的介质中，则只利用葡萄糖，而乳糖操纵子蛋白质处于低水平。同样如果既有萄葡糖又有半乳糖存在时，半乳糖操纵子

即无活性。葡萄糖比其它糖类被优先利用的原因还不清楚。这可能是由于在进化过程中细菌经常处于富含葡萄糖的环境之中，长期适应的结果。

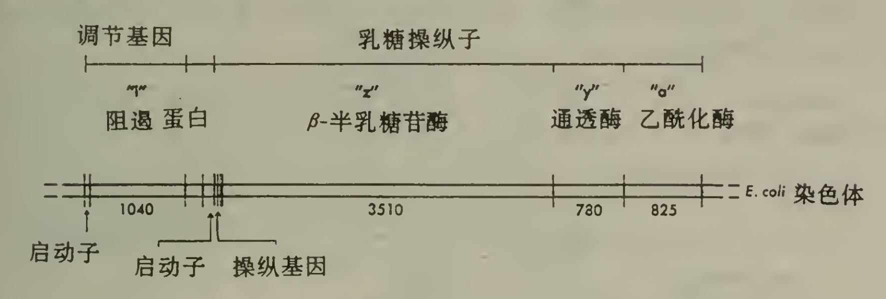


图 10-7 乳糖操纵子及其有关的调节基因

数字表示每个基因的碱基对数。

葡萄糖的抑制作用并非由于不同糖类进入细胞的速度有所不同，而是由于转录作用受到影响。这首先是由于发现了使乳糖操纵子对葡萄糖不敏感的控制区发生了突变的结果。用这种突变体乳糖操纵子可以最大限度地诱导，即使有大量葡萄糖存在之时也是如此。

(一)阻遏物

RNA 聚合酶(与启动子结合)及阻遏蛋白(与操纵基因结合)与 DNA 结合部位的阐明向我们指出阻遏蛋白是如何工作的。它们以物理方式阻止在启动子上 RNA 聚合酶的结合以及 RNA 合成的起始。乳糖操纵子是由同样的 21 个 DNA 碱基对组成的，它们是乳糖操纵子 mR-NA 最初 21 个核苷酸的模板。因此，lac 阻遏物覆盖着模板，恰恰是 RNA 聚合酶要占据的催化中心，从而阻止它靠近模板。其它操纵基因都与启动子的不同部位重叠，但它们都干扰 RNA 聚合酶的结合。恰如我们所预计，阻遏蛋白仅阻止 RNA 合成的起始，一旦延长开始后，对 RNA 链的生长就没有影响。

(二)辅阻遏物

阻遏物并非永远能阻止 mRNA 的合成。否则，它们就将永远抑制其特异蛋白质的合成。许多阻遏物分子能以活性的及无活性的两种形式存在，这要看它们与其适当的诱导物或辅阻遏物(corepressor)是否结合而定，诱导物的结合可使阻遏物失活。例如，当与一 β-半乳糖苷如乳糖或异乳糖(allolactose，乳糖的代谢物，为天然诱导物)结合时，lac 阻遏物即不能与其专一的操纵基因结合。因此，加 β-半乳糖苷于生长细胞中，以降低 lac 阻遏蛋白的分子浓度，可使 β-半乳糖苷酶得以合成。反之，辅阻遏物的结合则将无活性的阻遏物变为有活性的形式。例如，在细胞中加入色氨酸可以激活阻遏物，后者控制色氨酸生物合成所需酶的合成，这就迅速切断其专一的 mRNA 分子的合成(图 10-8)。

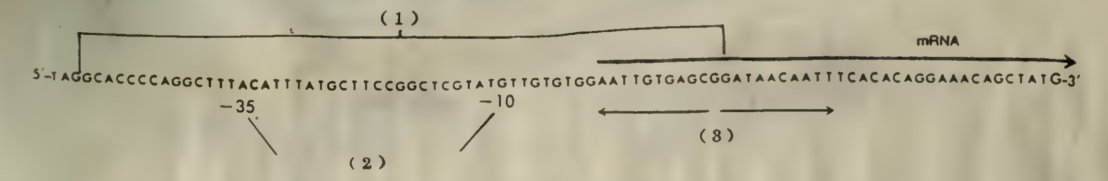


图 10-8 乳糖操纵子的启动子与操纵基因部位的重叠 DNA 链与 mRNA 在序列上一致

(1)DNA 阻遏蛋白结合区；(2)启动子同源识别系列；(3)对称的操纵基因。

——示被阻遏蛋白覆盖的启动子。

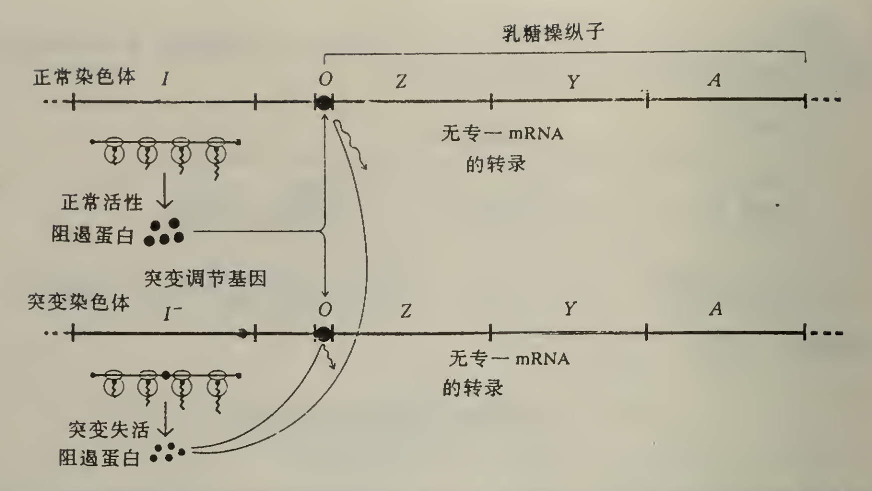
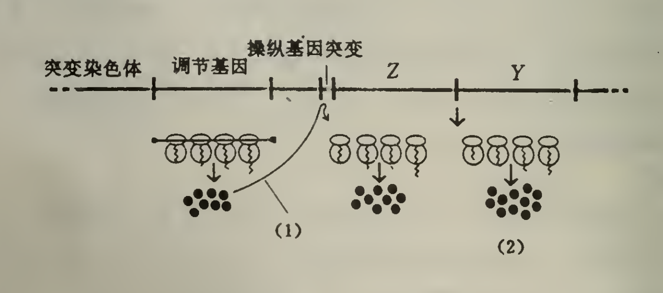


图 10-9　有功能阻遏蛋白比失活的阻遏蛋白占优势，在无外加 β-半乳糖苷时，β-半乳糖苷酶分子不产生

在阻遏物与其诱导物或辅阻遏物之间不形成共价键。每个阻遏物分子的一部分在形状上与其诱导物或辅阻遏物的专一部分是互补的。二者以次级键(氢键、盐桥或范德华力)相连结，由于这些键是弱键，它们可以快速形成或快速破坏，使阻遏物(活性的或非活性的)迅速调整，以适应生理的需要。例如，β-半乳糖苷酶 mRNA 的合成在去掉乳糖后几乎可以立即停止(图 10-9)。阻遏物与诱导物相结合构型上的可逆变化是别构酶转化的又一个例证(图 10-10)。

(a)



(b)

突变染色体
O^c
Z
Y
操纵子产物合成
正常染色体
O
Z
Y
β-半乳糖苷酶 mRNA 不合成

图 10-10　正常及突变的操纵基因对专一的 mRNA 合成的调控

(a)含突变操纵基因的细胞；(b)含正常操纵基因及突变操纵基因的细胞 O^c 比 O 占优势。
(1)阻遏蛋白不能与突变操纵基因结合；(2)在缺少诱导物时 β-半乳糖苷酶及通透酶之合成。

二、环 式 AMP

葡萄糖对转录的作用并不是直接的，其降解产物(降解物 catabolite)是靠降低细胞内的环式 AMP(cAMP)的含量起作用。此关键代谢物(cAMP)对葡萄糖降解代谢所抑制的各种操纵子的转录都是需要的，葡萄糖降解物控制细胞内 cAMP 水平的方式还不清楚。ATP 是 cAMP 的直接代谢前体，担任这种转化的酶是腺苷酸环化酶(adenylcyclase)，此酶可能受代谢降解物的直接抑制。另外有一种磷酸二酯酶(phosphodiesterase)，能将 cAMP 转变成 AMP，故这种抑制作用可能受 cAMP 分解的速度所控制(图 10-11)。

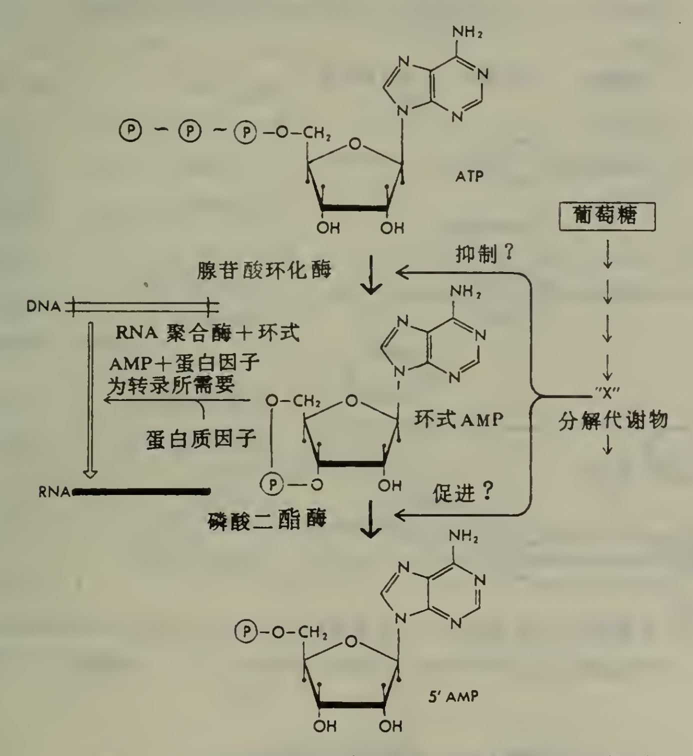


图 10-11　环式 AMP 对代谢降解物敏感的转录的控制

环式 AMP 并不直接促进 lac mRNA 的合成，而靠结合在代谢降解物基因激活蛋白(catabolite gene activator protein，简称 CAP)上起作用。CAP 是一个分子量为 45 kD 的二聚体。它对转录没有影响，只有 cAMP 与它结合才起作用。此后它就获得与 DNA 专一部位结合的能力，这时就增加邻近操纵子的转录速度。CAP 对一切葡萄糖敏感的操纵子都是正调控因子，故突变的细胞都不能利用大多数的糖类。

三、CAP 对 RNA 聚合酶的控制

CAP 能控制 RNA 聚合酶在乳糖启动子上的结合。CAP 或者阻遏物对 mRNA 生长速度都没有任何影响。但二者则控制 RNA 聚合酶分子与启动子结合的速度，一个为正调控，另一为负调控。实际上阻遏物阻断 RNA 聚合酶的结合，CAP 帮助 RNA 聚合酶有效地与乳糖启动子结合，使更多的 RNA 的合成起始进行。有如操纵基因一样，CAP 与乳糖操纵子 DNA 结合的部位也由阻止其功能的突变做出了标记，即在 cAMP 较高的条件下并不刺激乳糖酶，用足

迹法鉴定出一个专一的结合序列，重要的结果是 CAP 及 cAMP 的复合物在启动子上与 RNA 聚合酶相结合（见图 10-12 及图 10-13）。

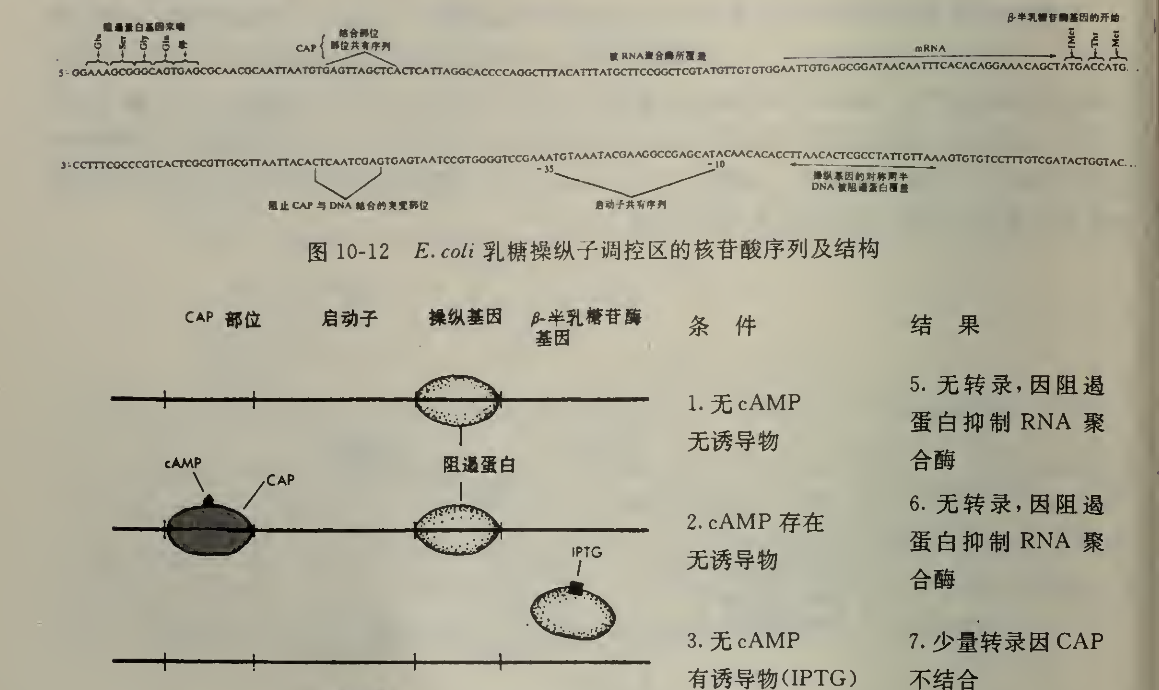


图 10-12　*E. coli* 乳糖操纵子调控区的核苷酸序列及结构

图 10-13　乳糖操纵子转录被 CAP、阻遏蛋白、cAMP 及诱导物（如天然诱导物异乳糖或合成的诱导物，异丙基硫半乳糖苷 IPTG）的调控

第七节　阿拉伯糖操纵子

细菌中有一个阿拉伯糖操纵子（arabinose operon），在葡萄糖馈乏时，它能利用阿拉伯糖作为能源。阿拉伯糖 *C* 基因编码一种 C 蛋白，它能激活阿拉伯糖 BAD 操纵子的转录，此操纵子由 *araB*，*araA* 及 *araD* 基因组成（图 10-14）。

C 蛋白必须与阿拉伯糖形成复合物，才能完成调控，因此只有当底物存在时，酶才能生成。与乳糖操纵子相似，阿拉伯糖操纵子对葡萄糖很敏感，因此除阿拉伯糖及 C 蛋白外，它也被 AMP 及 CAP 控制。两种激活剂，C 蛋白及 CAP 如何调节一个操纵子呢？

这个机理比较复杂，但是可以理解。C 蛋白既是激活剂又是阻遏物，而且它以两种不同方式作为阻遏物。C 蛋白在阿拉伯糖操纵子中的 3 个不同部位上结合，控制 araI，$araO_1$ 及 $araO_2$ 区（图 10-15）。

中间区 $araO_1$ 与 *C* 基因的启动子重叠，当细胞中有足够的 C 蛋白积累时，只起阻遏 *C* 基因 mRNA 合成的作用。因此*C*基因是自我调节的（autoregulated）。另外两个C蛋白结合部

位，araI 及 $araO_2$ 则与 CAP 结合部位共同起着调节 BAD 操纵子的阻遏及激活作用，据认为当 cAMP 浓度低，CAP 结合部位空缺时，一个 C 蛋白分子（可能是二聚体或四聚体）. 即与 araI 及远离的 $araO_2$ 部位结合，于是将 DNA 拉成一个环（见图 10-15）。在这个阻遏模型中，不论阿拉伯糖是否与之结合，C 蛋白都不能在 BAD 启动子上激活 RNA 聚合酶。然而 araI 后面 CAP-cAMP 的结合改变位于 araI 的 C 蛋白的活性，可能靠直接与之接触。这时 C 蛋白放出 $araO_2$ 部位，于是并不形成环，而它也能激活 RNA 聚合酶，以使邻近的 BAD 启动子开始合成，对C蛋白激活BAD转

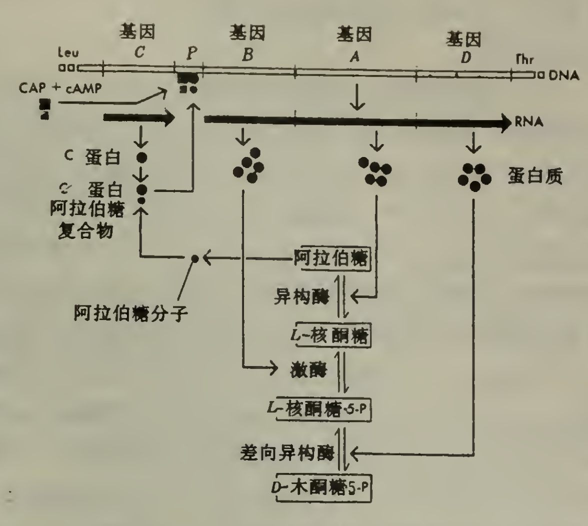


图 10-14 大肠杆菌的阿拉伯糖操纵子（正调控）

(a)阿拉伯糖操纵子的调节部位

C-蛋白结合部位
$araO_2$
$araO_1$
araI
启动子
启动子
CAP-蛋白结合部位
C mRNA
BADmRNA
148 核苷酸

RNA 聚合酶
C 蛋白
(=阿拉伯糖
CAP + 环式AMP

(b) 阿拉伯糖操纵子的控制

条件	结果	构型
1.	A	
2.	B	
3.	C	

1. $araO_2$ ara O_1 araI C mRNA
2. ara O_1 $araO_2$ araI
3. cAMP 阿拉伯糖 $araO_2$ ara O_1 araI BAD mRNA

图 10-15 大肠杆菌中阿拉伯糖操纵子的调节

(a)阿拉伯糖操纵子的调节部位；(b)阿拉伯糖操纵子的控制。1. 无 C 蛋白；2. C 蛋白存在，阿拉伯糖 cAMP 低；3. C 蛋白及阿拉伯糖。A. BAD 启动子不活化，C 基因转录；B. BAD 启动子阻遏，C 基因启动子阻遏；C. BAD 启动子活化，C 基因启动子阻遏。

录来说，它必须首先结合阿拉伯糖，所以，两个小分子，cAMP 和阿拉伯糖，对 BAD 有效的表达是必需的。

至于乳糖操纵子，细胞利用操纵子上的专一的调节因子(lac 阻遏物或 C 蛋白)和 CAP 蛋白去检测专一性糖的存在及葡萄糖的缺少，它是与阿拉伯糖操纵子不同的机制。像阿拉伯糖 C 蛋白那样，任何 DNA 结合蛋白可能既是激活剂又是阻遏物，这要看它是否阻断或促进 RNA 聚合酶与启动子结合而定。而噬菌体 λ 阻遏物则二者兼而有之。当葡萄糖受限制时，有时 CAP 蛋白也充作阻遏物，阻止不需要的基因的表达。*E. coli* 的乳糖操纵子中 CAP 蛋白，既作为激活剂又作为阻遏物的这种复杂控制即是一例。此外，还有一个结合两个操纵基因的半乳糖调节的阻遏物，这样可以将 DNA 拉成一个环，有如阿拉伯糖操纵子那样。最近的研究表明，甚至在乳糖操纵子上，CAP 可能也有一个像阻遏物的次要功能。除去乳糖操纵子的主要启动子之外，还有一个第二位强的启动子上游区，在该处 RNA 聚合酶强烈结合，但是并不有效地起始 RNA 的合成，因为这两个启动子重叠，这种结合可以阻断 RNA 聚合酶与主要启动子的相互作用。CAP 结合部位与第二个启动子重叠，因此 CAP 蛋白阻止 RNA 聚合酶结合在第二个启动子上，同时它直接激活 RNA 聚合酶结合在主要启动子上，CAP 这两种功能对 lac 操纵子的表达都是必要的。

CAP 及阿拉伯糖 C 蛋白代表两类细菌中存在的激活剂：一般激活剂如 CAP，其功能是在一定条件下所需要，如葡萄糖、氮或磷酸盐受限时；另一类是特异的激活剂，如阿拉伯糖 C，它们控制少数基因，其产物是将特殊化合物降解(如阿拉伯糖、麦芽糖或丝氨酸)。

第八节　氨基酸合成的操纵子

生物细胞中的氨基酸合成也是由操纵子调节的，细胞需要某种氨基酸时其基因即表达，不需要时基因即关闭，达到经济的原则。目前有多种氨基酸合成的操纵子的结构已研究清楚，如色氨酸、苏氨酸、组氨酸、苯丙氨酸、亮氨酸、异亮氨酸等十余种，现在我们详细讨论色氨酸操纵子(*trp*)的基因表达的调控。

一、色氨酸操纵子

色氨酸操纵子控制 5 种酶的合成，由 *trp E*,*D*,*C*,*B*,*A* 编码(图 10-16)。当色氨酸浓度低时，这 5 个邻近的 *trp* 基因才开始转录成 mRNA。色氨酸不是一个诱导物，而是作为一个辅阻遏物(corepressor)。辅阻遏物活化其专一的阻遏物，因此它能与 *trp* 操纵基因结合并阻止其相应基因的转录。当色氨酸浓度低时，*trp* 操纵基因不被占据，*trp* mRNA 的合成在邻近的启动子处开始进行。不过，*trp* mRNA 一旦起始合成，并不自动生长到全长度，而大多数 *trp* mRNA 分子在第一个 *trp* 基因(*trp E*)开始转录之前即停止生长(图 10-16)。接近此序列的一端及 *trpE* 之前为转录的终止子(terminator)，它由一个特征的 RNA 发夹环组成(图 10-16 中 3 区及 4 区的序列构成)，随后为 8 个尿嘧啶(U)残基，此处即所谓弱化子(attenuator)，RNA 合成通常在此即行停止，前导 RNA 长为 139 个核苷酸(图 10-17)，前导序列有 3 个特征可使 RNA 聚合酶通过弱化子。(1)除去由前导的 1 区与 2 区所形成的终止子发夹外还有第二个发夹；(2)2 区与 3 区互补，于是又形成一由 2 区及 3 区组成的另一发夹，从而阻止终止子发夹的形成；(3)引导 RNA 编码 14 个氨基酸组成的短前导肽，其前面为一长的核糖体结合位点。前导肽有一特殊的序列特征：即 2 个色氨酸密码子排成一列，在编码制造氨基酸的酶的其它操纵子的相应序列中

尚未发现(表 10-2)。而亮氨酸的前导肽有 4 个相邻的亮氨酸密码子排成一列,组氨酸操纵子前导肽有 7 个组氨酸密码子排成一列。

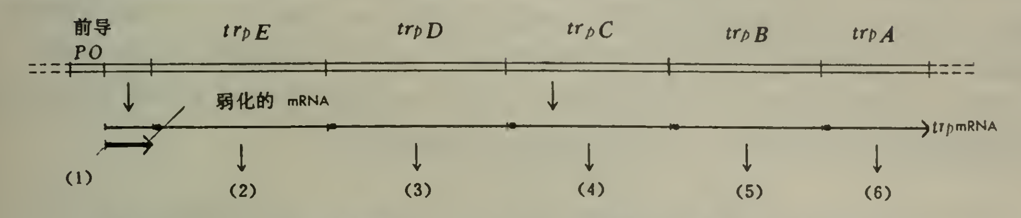


图 10-16 *E. coli* 的色氨酸操纵子,表示前导肽与编码 trp 酶的结构基因的关系

(1)核糖体结合部位;(2)氨基苯甲酸合成酶;(3)磷酸核糖氨基苯甲酸转移酶;(4)磷酸核糖氨基苯甲酸异构酶-吲哚甘油磷酸合成酶;(5)β-色氨酸合成酶;(6)α-色氨酸合成酶。

表 10-2 弱化子控制的含有氨基酸生物合成的操纵子的前导肽

操纵子	前导肽的氨基酸序列
色氨酸	Met Lys Ala Ile Phe Val Leu Lys Gly Trp Trp Arg Thr Ser
苏氨酸	Met Lys Arg Ile Ser Thr Thr Ile Thr Thr Thr Ile Thr Ile Thr Thr Gly Asn Gly Ala Gly
组氨酸	Met Thr Arg Val Gln Phe Lys His His His His His His His Pro Asp
异亮氨酸- 缬氨酸 GEDA	Met Thr Ala Leu leu Arg Val Ile Ser Leu Val Val Ile Ser Val Val Val Ile… …Ile Ile Pro Pro Cys Gly Ala Ala Leu Gly Arg Gly Lys Ala
亮氨酸	Met Ser His Ile Val Arg Phe Thr Gly Leu Leu Leu Leu Asn Ala Phe Ile Vel… …Arg Gly Arg Pro Val Gly Gly Ile Gln His
苯丙氨酸	Met Lys His Ile Pro Phe Phe Phe Ala Phe Phe Phe Thr Phe Pro
异亮氨酸-缬氨酸 B	Met Thr Thr Ser Met Leu Asn Ala Lys Leu Leu Pro Thr Ala Pro Ser Ala … … Ala Val Val Val Val Arg Val Val Val Val Val Gly Asn Ala Pro

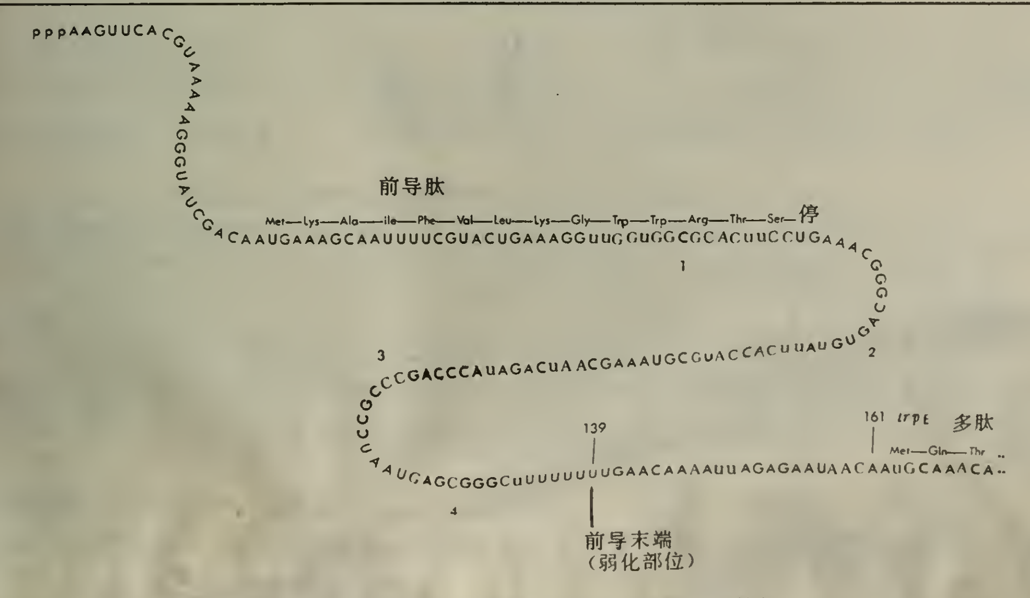


图 10-17 trp 操纵子的前导肽 RNA 核苷酸序列特征

这些密码子的功能是使企图翻译前导肽的核糖体停止前进。而当色氨酸缺乏时,核糖体达到色氨酸密码子时必须停止,围绕色氨酸密码子的 RNA 停下来结合在核糖体上,不能成为发夹环的一部分,图 10-18 即显示这个结果:在色氨酸密码子处被捕捉的核糖体被放在可使 2 区

与 3 区自由配对的位置上，于是终止子发夹(3,4)不能形成，RNA 聚合酶跨过弱化子达到操纵子，使 trp 酶表达。反之，如果有充分的色氨酸，使核糖体得以通过色氨酸密码子前进，则核糖体阻止序列 2，使 3，4 区形成终止子发夹，则转录在前导 RNA 末端流产。第三种可能性是在色氨酸密码子之前翻译停止(因为某些其它氨基酸缺乏)，或者不能使翻译起始进行，那么，序列 1 与序列 2 自由配对，则终止子结构再度形成，以停止 trp 操纵子的表达。前导肽被认为在合成时对核糖体的就位是重要的，但立即就被细胞的蛋白酶破坏掉。

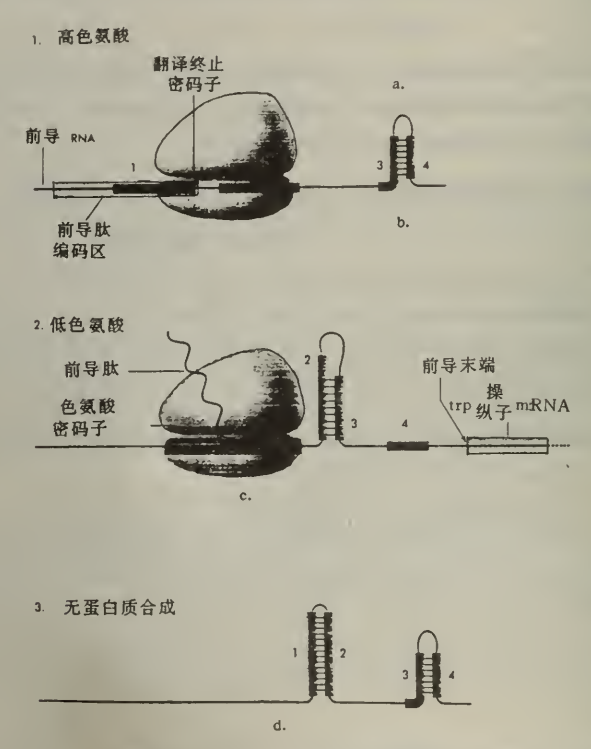


图 10-18 trp 操纵子的弱化子上的转录终止由色氨酸的供应控制

a. 序列 3 与序列 4 配对，形成转录终止发夹，RNA 合成停止，trp 酶不表达；b. 核糖体前进至前导肽终止密码子，封闭序列 2；c. 核糖体在邻近的 trp 密码子停在让序列 2 与 3 配对，阻止 3，4 终止发夹形成，RNA 聚合酶进入操纵子，制造 trp 酶的 mRNA；d. 如核糖体在开始前导肽 AUG 的翻译，1，2 发夹形成，阻止 2，3 形成。

阻遏作用与弱化子控制之间可能同时存在，使得细胞内色氨酸微细调节到一个水平，使得两步反应更紧急地达到色氨酸饥饿，起始反应是停止阻遏物的结合，更大的饥饿则导致弱化作用的松弛，其它氨基酸的操纵子、如组氨酸及亮氨酸操纵子没有阻遏物，它们完全依靠弱化作用来控制。

下面再讨论一下弱化子的重要性。色氨酸操纵子的转录必须由一个可控制的终止部位来调节，这个部位就是弱化子。它位于操纵子与这个氨基酸合成途径的第一个酶的基因之间，如图 10-19。

弱化子这个调节部位与某些操纵子末端的停止信号相似，含有一个富含 GC 的序列，随后为一个富含 AT 的序列。这两个区域都以弱化子为中心，表现出对称的结构，如图 10-20。

弱化子部位与调节 *trp* 基因转录的操纵子部位是互补的，当色氨酸充裕时，色氨酸转录的起始就被色氨酸阻遏物复合物与操纵子相结合而阻拦住；而当细胞中色氨酸浓度下降时，阻遏

就被取消，转录又再开始运行。

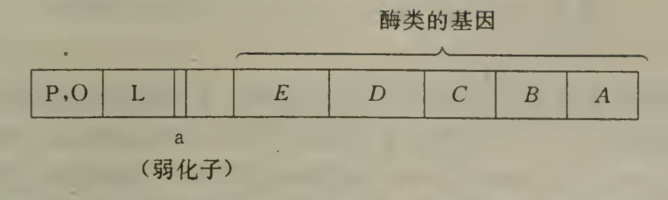


图 10-19　trp 操纵子的示意图

P 为启动子，O 操纵子，a 弱化子，L 为前导肽，E，D，C，B，A 为色氨酸合成途径中的酶类基因，在 trp 操纵子中操纵子(O)与启动子(P)是重叠的。

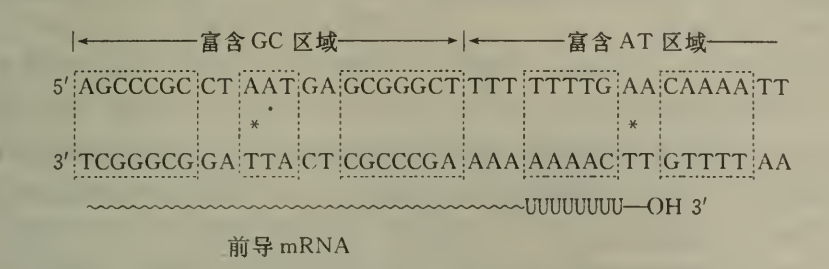


图 10-20　trp 弱化子部位的碱基序列

富含 GC 区域与富含 AT 区域中的碱基部都具有二重轴对称的结构。

二、组氨酸等的操纵子

大肠杆菌其它氨基酸合成的操纵子也都含有弱化子。每个操纵子的前导肽(leader peptide)链中含有该操纵子所调控的氨基酸的大量残基。例如，苏氨酸操纵子含有合成苏氨酸及异亮氨酸的酶，其前导肽在 16 个残基序列中含有 8 个苏氨酸及 4 个异亮氨酸，如图 10-21。

A.　Met- Lys- Arg- Ile- Ser- Thr- Thr- Ile- Thr- Thr- Thr- Ile- Thr- Ile- Thr- Thr

（苏氨酸 5′ AUG AAA CGC AUU AGC ACC ACC AUU ACC ACC ACC AUC ACC AUU ACC ACC 3′ 操纵子）

B.　Met- Lys- His- Ile- Pro- Phe- Phe- Phe- Ala- Phe- Phe- Phe- Thr- Phe- Pro- Stop

（苯丙氨酸 5′ AUG AAA CAC AUA CCG UUU UUC UUC GCA UUC UUU UUU ACC UUC CCC UGA 3′ 操纵子）

C.　Met- Thr- Arg- Val- Gln- Phe- Lys- His- His- His- His- His- His- His- Pro- Asp

（组氨酸 5′ AUG ACA CGC GUU CAA UUU AAA CAC CAC CAU CAU CAC CAU CAU CCU GAC 3′ 操纵子）

图 10-21　mRNA 相应部分的前导肽的氨基酸序列及碱基序列

A. 苏氨酸操纵子；B. 苯丙氨酸操纵子；C. 组氨酸操纵子。

在苯丙氨酸操纵子的前导肽的 15 个残基序列中存在 7 个苯丙氨酸，而在组氨酸操纵子的前导肽所含的 7 个组氨酸残基则列成一排。很明显，这些前导肽的 mRNA 正是设计成能够察觉出所被合成的蛋白质的氨基酸水平。如果相应的氨基酰 tRNA 缺乏时，前导肽的翻译就自动停止下来。例如，在色氨酸操纵子中被阻止的核糖体促使 mRNA 的构象开放，使 RNA 聚合酶能阅读过弱化子部位。在组氨酸操纵子前导 RNA 存在 7 个编码组氨酸的密码子就强化了检查系统的灵敏性。这个组氨酸操纵子可以单纯靠弱化子的结构实现它的控制作用。

第九节　基因表达翻译水平的调节

原核生物中基因的表达也在翻译水平上受到控制。蛋白质翻译的起始主要依靠AUG起始密码子上游富含嘌呤的6～8个核苷酸之存在与否。1974年Shine与Dalgarno首先指出这个核糖体结合位点(ribosome-binding site)的存在，这段DNA序列就称为Shine-Dalgarno序列。他们观察到在细菌核糖体中的16 S小亚基rRNA3′端在序列上与核糖体结合位点序列是互补的。mRNA与核糖体结合序列一般集中在起始密码子上游的8个核苷酸上。如果这个区域发生突变，就会大大降低其相应的mRNA的翻译效率，不论这个突变靠近AUG密码子，或远离密码子都是一样。这样的序列被包在环卡结构(loop structure)之中，因而无法与16 S rRNA相互作用，进行翻译。

近来有证据表明Shine-Dalgarno序列的工作效率受与它们结合的蛋白质控制，从而阻碍其作用。*E. coli*的核糖体蛋白(ribosomal protein，即γ-蛋白)就是一个例证。当γ-蛋白合成的速度超过γ-RNA合成的速度时，游离的γ-蛋白就积累起来，有一些"关键的γ-蛋白"就与Shine-Dalgarno序列结合。这样的话，核糖体蛋白就不能比它们用于制造核糖体时合成得更快。

γ-蛋白不同组合的基因包含在*E. coil*基因组一些不同的操纵子上，每个操纵子编码其自己的关键蛋白质，它抑制整个操纵子的表达。在核糖体组装时，关键的γ-蛋白与γ-RNA结合得早些。与这些蛋白质结合的γ-RNA的序列及与关键蛋白质相互作用的γ-蛋白的mRNA的序列是十分相似的。因此，翻译水平的调控就是γ-RNA与γ-蛋白mRNA之间相互与其关键蛋白结合的竞争。当γ-蛋白mRNA被所结合的关键蛋白使它不能结合时，它们就比平时更快地降解。蛋白质翻译一旦开始后，其速度即受对应于mRNA分子中所利用的各种tRNA的供应情况而决定。在细胞中不同tRNA分子的数量是不同的。大量存在的那些tRNA通常是对应最常使用的密码子，因此，被稀少的tRNA分子所识别的密码子大都翻译得较慢，而被丰富的tRNA识别的密码子则翻译得要快些。

担任以mRNA为模板翻译成蛋白质的核糖体是一个很大的生物大分子，含有50多种核糖体蛋白及许多种核糖体RNA(rRNA)。在蛋白质合成时，这50多种蛋白质如何协调工作是一个非常复杂的问题。它们的基因分别定位于20多个操纵子上，它们的合成是精确而平衡的。因此，严格的调控是非常必要的。核糖体成分构成了细菌细胞总干重的40%左右。核糖体蛋白合成的调控主要是在翻译水平上，而不是在转录水平上完成的，这一点与大多数其它蛋白质不同。例如，至少每种核糖体蛋白的操纵子都有一种蛋白质当作它的翻译阻遏蛋白(translational repressor)，它定位在靠近合成的起始部位的mRNA上，从而阻止许多种蛋白质的合成(图10-22)。

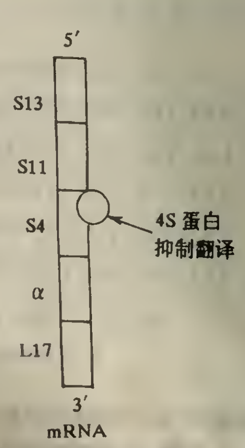


图10-22　翻译阻遏的反馈调控

一个核糖体蛋白(4 S)阻遏编码核糖体大(L)、小(S)亚基及RNA聚合酶α亚基的翻译过程。

这种与mRNA结合并不影响核糖体的聚合，这是因为核糖体蛋白与核糖体RNA的结合比其与mRNA的结合更紧。只有当核糖体蛋白的生产超过核糖体RNA时，翻译的阻遏才会发生，或者当这些核糖体蛋白的形成不是以等摩尔生成时翻译的阻遏也会发生。

核糖体RNA及转移RNA(tRNA)的合成与非核糖体蛋白的合成也是相互协调的。当氨基酸缺乏时，这些RNA即停止合成。当氨基酸缺乏时从ATP和GDP合成一种新的化合物，称为5′-二磷酸-乌苷-3′-二磷酸(5′-diphosphate-gunosine-3′-diphosphate，ppGpp)，从ATP分子上转移2个磷酸分子到GDP的3′-羟基(—OH)上，ppGpp也可以从ATP及GTP形成的pppGpp经过水解而生成ppGpp。当核糖体的A部位上存在未氨酰化的tRNA时，ppGpp的合成即发生。因此，ppGpp和pppGpp只在缺乏氨基酸的细胞中形成。这是由于在不饥饿的细胞中A部位总是由氨酰tRNA占据着，一旦ppGpp形成，它即抑制操纵子转录的起始，并使许多其它转录本(transcript)的延长变得缓慢下来。

参考文献

1. Stryer. L. Biochemistry (4th ed.). W. H. Freeman and Company, New York: 1995.
2. Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. and Weiner, A. M. Molecular Biology of the Gene. Vol. I. (4th ed.). Benjamin/Cummings, Menlo Park: 1987.
3. Lewin, I. V. Genes V. Oxford University Press, London: 1995.
4. Freifeldar, D. Molecular Biology (2nd ed.). Johns and Bartlett Publisher, Boston: 1987.

(阎隆飞)

第十一章　病　　毒

第一节　引　　言

由于病毒的寄生涉及各种不同类型的生物，所以病毒之间也存在着极大的差异。病毒在生物学上的分化程度，比细菌、植物以及动物各类群的总和还要高。对于这种多样性的理解、认识过程，构成了对病毒本身以及病毒与相应寄主相互作用关系的研究历史。蛋白与蛋白、蛋白与核酸、蛋白与脂类间的相互作用决定了病毒粒子(virions)的结构、病毒基因组的合成与表达以及病毒对寄主细胞的影响。这便是在分子水平上的病毒学。

一、病毒的定义

简单地说，病毒就是一类亚显微专性细胞内寄生物。虽然这一定义可以将病毒区别于其它各类高等的或“可见”的生命有机体，但它又与少数具有特殊寄生性生活史的原核生物相混淆。如立克次氏体和衣原体，它们均是专性细胞内寄生细菌，仅能在寄主细胞外存活很短一段时间或保持侵染性。因此，以下内容以对病毒的定义也是必要的：(1) 病毒粒子(virions)由蛋白质和核酸构成，其中的核酸仅为DNA 或RNA 中的一种；(2) 病毒粒子是由预先形成的组分装配而成的，自身不进行“生长”或发生分裂。其它生命形式则是通过其各种组分综合数量的增加而“生长”，并通过分裂进行繁殖；(3) 病毒自身不具备进行能量代谢和产生蛋白质合成所需核糖体的遗传信息，在能量代谢上绝对依赖于寄主细胞。

与其它生物类型相比，另外一些病原体具有与病毒更加相近的特性。如：类病毒(viroids)、拟病毒(virusoids，目前多称作与类病毒相似的卫星RNA-viroidlike satellite RNA)和侵染性蛋白(prions)。类病毒是非常小的环状RNA 分子(200～400 nt*)，并具有棒状二级结构。它们没有外壳蛋白或包膜，但与某些植物病害相关。其复制策略与病毒相似，也是细胞内专性寄生物。拟病毒是分子量较大(约1 000 nt)，是与类病毒相似的“卫星”式分子，它依赖于病毒的复制而增殖，并被包被于病毒的衣壳蛋白中。侵染性蛋白是一种定义比较模糊的侵染物质，它由单一种类的蛋白分子构成，而不含有核酸。这些蛋白多与“慢性病毒病”相关，如人类C-J 病、羊搔痒病及牛海绵状脑炎(疯牛病)。

二、病毒学的起源及发展

1892 年，俄国植物学家D. Iwanowski 证明，感病烟草的汁液用当时孔径最小的细菌滤器过滤后，仍可将病害传染给其它植物。稍后的1898 年，M. Beijerinick 充实了Iwanowski 的结果，确认了烟草花叶病毒的存在，成为现代病毒学的创始人。他将病毒称为“感染性活

---

* nt 表示核苷酸数目。

液”(contagium vivum fluidum)。此后 20 年里，世界各国的科学家分别在牛的口蹄疫、人类脊髓灰质炎以及细菌中发现了不同类型的病毒，其中侵染细菌的病毒称为噬菌体(bacteriophages)。自 30 年代以来，许多人对病毒学的不同方面进行了研究，其中包括病毒的结构、遗传和复制等。同时，病毒学的历史也是实验设备及系统的发展过程。如：动植物活体寄主体系、细胞培养方法、血清及免疫学技术、超微结构研究以及大量的现代分子生物学技术。利用这些技术及系统，不仅可以对病毒进行测定，也成为生物学的一个新的领域。其中不仅包括病毒本身的生物学，而且也包括了病毒赖以生存的寄主细胞生物学。

回顾现代分子生物学的发展过程，我们随处可见病毒学所发挥的重要作用，许多突破性的进展都是来自于病毒学的研究结果。例如，1935 年，Stanley 以烟草花叶病毒(TMV)为材料，首次获得蛋白质结晶，并进行了 X 光衍射分析；1952 年，Hershey 和 Chase 利用 T2 噬菌体对大肠杆菌的感染，证明 DNA 含有遗传信息；1967 年，Sinheimer 根据对单链 DNA 噬菌体 $\Phi X174$ 的研究结果，提出了基因的“滚环式”复制模型；1970 年，Baltimore 和 Temin 在纯化的罗斯肉瘤病毒中发现了逆转录酶，证明可以由 RNA 反向产生 DNA，使“中心法则”得到了完善和补充；1971 年，Schneider 在烟草环斑病毒的卫星 RNA 中，发现了具有酶活性的核酸片段，核酶(ribozyme)；1977 年，Sanger 以 $\Phi X174$ 噬菌体为材料，建立了末端终止 DNA 序列分析方法。这些成就不仅极大的丰富了分子生物学的内容，也体现了病毒作为一种生物类群，在遗传构成、基因表达方式以及与其它生物相互关系上的特殊性。病毒基因组一般要比真核生物基因组小得多，而且便于大量纯化获得，更重要的是因为一些较小的真核病毒基本上是依靠寄主所具有的系统进行复制、表达和调控的，因此，对于病毒基因组遗传规律的研究结果，在一定程度上也就反映寄主基因组本身的一些规律，成为我们认识其它生物类群的一个窗口。

随着对于病毒本质和遗传规律研究的深入，以及相关分子生物学理论、技术的发展，病毒学研究在利用病毒开展基因工程方面也有了迅速的发展。例如，动植物病毒和噬菌体基因组及其启动子等调控元件，已经被广泛用于构建不同类型的基因工程载体，在外源基因对细菌和真核细胞的转化及表达调控中发挥着重要作用，如 λ 噬菌体、SV40、花椰菜花叶病毒等。另外，在病毒病的预防、治疗中，具有广谱性抗病毒活性的多肽——干扰素，在国内外均有基因工程产品的开发应用；乙型肝炎等病毒的基因工程疫苗也在病毒性疾病的预防中得到了广泛使用。继 1986 年 R. Beachy 首次利用烟草花叶病毒外壳蛋白基因，转化获得对病毒侵染具有抵抗作用的基因工程烟草植株以来，我国的科技工作者在利用病毒基因构建抗病毒转基因植物方面取得了许多成绩。已经用不同病毒基因或结构，转化获得了多种对病毒危害具有抗性的作物或品种，在抗病毒基因工程育种的应用方面居于世界领先水平。

第二节　病毒粒子的结构

一、病毒粒子的功能

最简单地来说，病毒粒子的外壳具有保护核酸基因组免受物理及化学因素和各种酶类破坏的功能。当离开寄主细胞时，病毒便处于一种不利其生存的环境中，无保护的基因组可能迅速失活。核酸对物理损伤和紫外线(如日光)造成的化学修饰非常敏感。由于自然环境中大量

存在的核酸酶，可以造成病毒基因组单链核酸中的磷酸二酯键断裂，或对核苷酸进行化学修饰，这些都足以使基因组不能复制，从而对病毒粒子产生抑制作用。病毒衣壳(capsid)中的蛋白亚基是大量存在的，也就是每个粒子中都含有许多拷贝。对一个或多个亚基所造成的损伤可能使特定的亚基丧失功能，但就整个粒子来说却仅造成对侵染性的有限破坏。因此可以说衣壳是一种有效的屏障。

另外，病毒的外部表面还与对寄主细胞的识别以及侵染初期的相互作用有关。虽然这个过程基本上是通过一类特异性的病毒附着蛋白与细胞受体的结合来实现的，但在侵染的起始阶段，衣壳也具有通过某种结构形式传递病毒基因组的能力，在这种结构中的基因组可以与寄主细胞相互作用。有时，这仅是一个将病毒基因组释放到细胞原生质体中的简单过程，但也可能是一个非常复杂的过程。例如，在逆转录病毒中，当基因组仍然位于病毒粒子内部时便会进行大量的修饰，在进入寄主细胞核之前由两条单链 RNA 分子转变成为一条双链 DNA 分子。因此，蛋白衣壳在病毒侵染过程的建立中发挥着生物学功能。

二、衣壳的对称性和病毒的结构

病毒的衣壳肯定是由多个蛋白分子，也就是亚基(subunit)结构所构成的。1955 年 Fraenkel-Conrat 和 Williams 证明，当用纯化的烟草花叶病毒(TMV)RNA 与外壳蛋白混合保温时，可以形成病毒粒子。这一发现表明病毒粒子的结构是处于最低的能级状态，也就是各种组分的最适结构形式。结构上的稳定性是病毒粒子的一个重要特征。有些病毒非常脆弱，不能在寄主细胞外存活，但许多病毒却能存活较长一段时间，有时可达数年。

在病毒组装的过程中，所涉及的作用力包括疏水力和静电作用力，另外还有少量的共价键与多个亚基间的结合有关。这意味着蛋白与蛋白、蛋白与核酸以及蛋白与脂类间的相互作用。目前对大多数病毒结构中的这类相互作用的细节尚未充分认识，但是已经对其中的一些基本原则和结构型式有了一些了解，它们作用于各种不同类型的病毒结构。以下将要讨论的是病毒结构的两种主要类型：即螺旋对称和 20 面体对称。

(一)螺旋体衣壳

烟草花叶病毒(TMV)是螺旋体(helix)对称病毒结构的代表。它以旋转对称方式将多个相同的蛋白亚基有序地组合到一起，也就是将不规则外形的蛋白质沿环形圆周排列，形成一个盘状结构。多个盘状结构相互堆集形成一个柱状体，而病毒基因组则被蛋白外壳包含在中空的柱状体内部。

X 光衍射分析表明，TMV 粒子的衣壳实际上是一种螺旋体结构。它包括了直径和螺距两个较为稳定的数学参数(图 11-1)，也是一种较为简单的结构。在 TMV 的粒子中，构成螺旋体一个圆周的亚基数量(μ)为 16.3，每个亚基轴向上升的高度(p)为

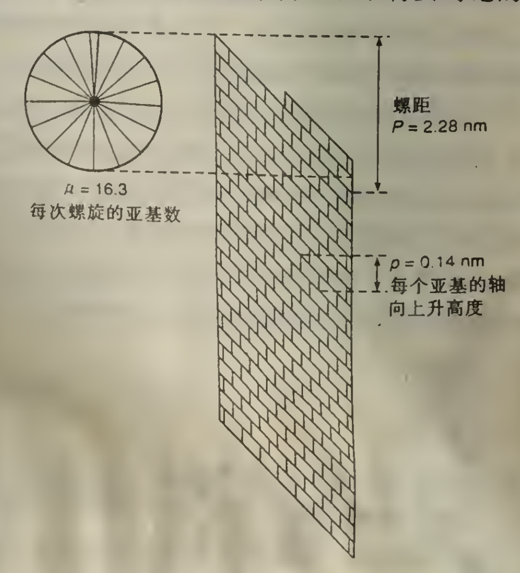


图 11-1　由单一种类的多个外壳蛋白稳定排列构成的 TMV 衣壳是一种螺距为 2.28 nm 的螺旋体

0.14 nm，而螺旋体的螺距(P)则等于 $16.3 \times 0.14 = 2.28$ nm。也就是 $P=\mu\times p$。TMV 的粒子为刚直棒状结构，有些螺旋体病毒则具有一定的柔韧性，较长的螺旋体病毒粒子一般都表现为弯曲状。

单一种类的蛋白亚基通过螺旋对称形成病毒粒子，已经在许多不同种类的病毒中得到证实。丝状噬菌体科中的一些噬菌体便具有最简单的螺旋状衣壳，如 M13 和 fd(即 Ff)噬菌体。这些噬菌体长度约 900 nm，直径约 9 nm，含有 5 种蛋白(图 11-2)。主要外壳蛋白是噬菌体 g8p，在每个粒子中，这种蛋白的拷贝数为 2 700～3 000。同时，在线状粒子的两端还存在其它 4 种其它的蛋白(g3p，g6p，g7p 和 g9p)，它们的拷贝数各约为 5 个。成熟的 g8p 分子大约由 50 个氨基酸残基构成，几乎完全是一种由 α-螺旋构成的短棒状结构。在短棒状结构中存在 3 个不同的区域：(1) 氨基末端的负电荷区，这一区域内含有酸性氨基酸残基并构成了病毒粒子外部的亲水表面；(2) 羧基末端的碱性正电荷区，它位于蛋白柱状体内侧并与带有正电荷的 DNA 基因组相邻；(3) 在以上两个区域之间还存在一个疏水区，它与 g8p 亚基之间相互作用、形成或稳定噬菌体粒子有关(图 11-2)。也就是说，Ff 噬菌体粒子是由外壳蛋白亚基间的疏水作用而形成的。螺旋体中连续旋转的 g8p 亚基与下部的亚基相互连接，与粒子长轴呈约 20°倾角并呈鱼鳞状互相重叠。它的 μ 值为 4.5，p 值为 1.5 nm。

由于噬菌体 DNA 包装在螺旋状粒子的中心，所以基因组的长度决定了粒子的长度。在 Ff 噬菌体制备物中都存在着各种不同的粒子，如：含有大于一个基因组 DNA 长度的多聚噬菌体(polyphage)、含有约 0.2～0.5个基因组长度缺失型 DNA 的小噬菌体(miniphage)以及含有长度大于一个基因组，但具有遗传缺损 DNA 的特长噬菌体(maxiphage)等各种形式。M13 线状粒子的这种特性已经被用于将其基因组改造成为一类克隆载体。外源 DNA 插入基因组所产生的重组噬菌体粒子，长度大于野生型粒子，但基因组能够被粒子包装的长度限制不明确。随着 M13 基因组长度的增加，其复制效率随之下降。当重组噬菌体基因组较野生型长度增加 1%～10% 时，无明显的不利表现；当长度较野生型增加 10%～50%时，复制明显减慢；而当增加长度超过正常基因组长度的 50%时，分离获得重组噬菌体则更加困难。

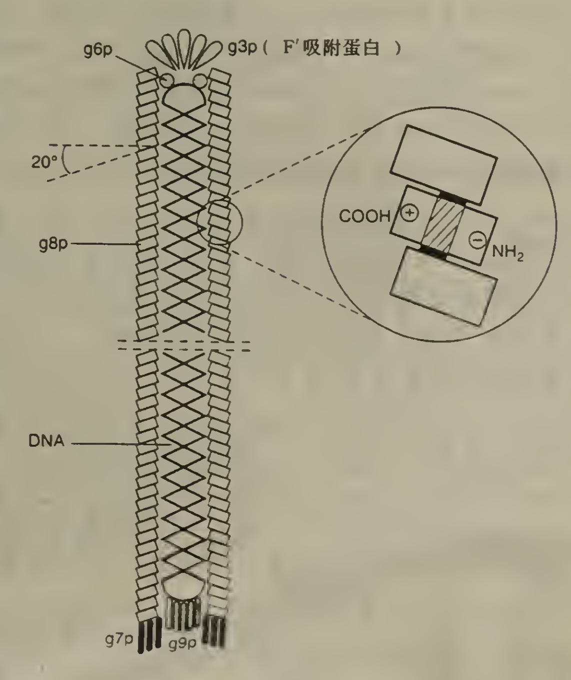


图 11-2　M13 噬菌体粒子结构示意图

衣壳结构在 Ff 噬菌体对适当宿主细胞进行侵染的过程中也具有作用。Ff 噬菌体为雄性宿主专一性噬菌体，在侵染过程中要求大肠杆菌表面的 F 菌毛。侵染过程的第一个步骤是，位于丝状粒子末端的 g3p 蛋白与 F 菌毛的末端发生相互作用，从而导致 g8p 蛋白的形态变化。g8p 蛋白的结构首先由 100%的 α-螺旋变为 85%的 α-螺旋，丝状粒子变短，附着于 F 菌毛上的粒子末端张开，噬菌体 DNA 外露。然后，g8p 蛋白亚基发生第二次形态变化，α-螺旋的比例由

85%下降为 50%，导致噬菌体粒子成为一个直径约 40 nm 的中空球体，噬菌体 DNA 排出，开始对寄主细胞的侵染。

许多植物病毒属于螺旋体对称，如烟草脆裂病毒组、烟草花叶病毒组、菌传棒状病毒组、大麦病毒组、马铃薯 X 病毒组、香石竹病毒组、马铃薯 Y 病毒组和线状病毒组。这些病毒的粒子长度由大约 100 nm（烟草脆裂病毒组）至 1 000 nm（线状病毒组）不等。植物病毒组所具有的这种结构可能与寄主植物细胞的生物学有关，或是与在寄主之间的传播方式有关。

相当数量的动物病毒也都属于螺旋体对称，但它们的外层另外还都含有一个脂类包膜(envelope)。这一类群中包括许多熟知的人类病原物。如：流感病毒（正粘病毒科）、腮腺炎病毒和麻疹病毒（副粘病毒科）、以及狂犬病毒（弹状病毒科）。所有这些病毒均为单链负意 RNA 基因组，而且分子结构相似。在被侵染细胞内，病毒核酸与一种碱性的核酸结合成蛋白聚集，构成螺旋状核衣壳(nucleocapsid)。这种蛋白质与 RNA 形成的复合体，除具有保护病毒基因组免受物理及化学因素伤害的功能外，有时还在病毒复制中发挥作用。

有些具有包膜的螺旋状动物病毒结构比较简单，如：狂犬病毒及非常相近的水泡性口炎病毒（VSV）（图 11-3）。它们为弹状病毒属成员，粒子沿着长约 11 000个核苷酸的负意 RNA 基因组进行组装。病毒 RNA 与碱性核衣壳(N)蛋白相互作用，形成一种螺距约 5 nm 的螺旋结构。这种结构在另外两种非结构蛋白（L 和 NS）的参与下形成病毒粒子的髓心。髓心核蛋白螺旋体由 30～35 次旋转构成，长约 180 nm，直径约 80 nm。单体 N 蛋白体积约为 9×5×3 nm，每个单体约可覆盖 RNA 基因组中的 9 个核苷酸。在上面谈到的丝状噬菌体粒子中，N 蛋白的作用是稳定 RNA 基因组，并保护其不受化学、物理及酶的破坏。对大多数具有包膜的病毒而言，核衣壳是由一种不定形的层体所包被，这一层体即可以与髓心作用，又可以与外部连接的类脂膜相互作用。这一层体被称为基质(matrx)。基质(M)蛋白在病毒粒子中通常为数量最大的蛋白质。例如：在 VSV 粒子中存在约 1 800 个基质蛋白拷贝和约 1 250 个 N 蛋白拷贝。

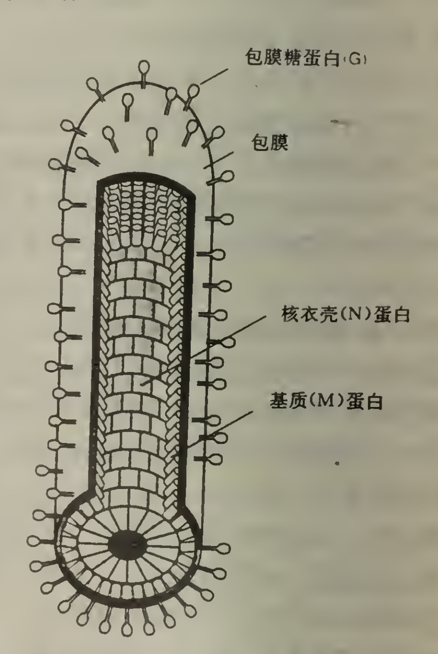


图 11-3 弹状病毒粒子结构示意图

目前已知，许多不同类型的病毒都是以螺旋对称方式存在的。基因组较小、结构简单的病毒，利用这种结构对基因组提供保护，而又不需要编码大量的衣壳蛋白(coat protein)。更为复杂的病毒则利用这种结构作为病毒粒子的基础，并以蛋白质和脂类表层进一步加工。

（二）20 面体（等轴体）衣壳

病毒衣壳结构的另外一种方式是蛋白亚基组装成一种类似于球形的结构，并将基因组包含在里面。蛋白亚基在立体表面的组装是一个比构成螺旋体更为复杂的过程。20 面体(icosahedron)是由 20 个三角形沿一个球形表面排列而成的立体形状（图 11-4）。

由重复亚基构建衣壳，病毒必须具有如何排列这些亚基的规则。对一个 20 面体来说，这种规则便是立体旋转对称，也称作 2-3-5 对称。它的含义为：①轴线穿过每条边线中点可产生 2

次旋转对称；②轴线穿过每个表面的中心可产生 3 次旋转对称；③轴线穿过每个顶点可产生 5 次旋转对称。

对于最简单的 20 面体衣壳来说，每个三角形表面是由 3 个相同的亚基所构成的。也就是说，构建一个完整的衣壳至少需要 60 个相同的亚基。少数几种简单的病毒粒子是通过这种方式构建的，如微小噬菌体科的 $\Phi X174$ 噬菌体，最近通过 X 光衍射对它的完整结构进行了确认。但是，由于蛋白质分子的形状并不规则，所以由此而形成的并不是真正的等边三角形或严格几何学意义上的 20 面体。

研究结果表明，在大多数情况下 20 面体病毒衣壳含有 60 个以上的亚基。由于所有亚基都是等价相连的，也就是它们之间的空间距离相等，并处于最小自由能级状态，所以由 60 个相同亚基构成的规则 20 面体是一种非常稳定的结构。由于一个真正的规则 20 面体只能由 20 个亚基组成，对于含有 60 个以上亚基的 20 面体来说，这些亚基之间要进行真正的等价连接，完全对称地排列是不可能的。为了解释这一问题，1962 年 Caspar 和 Klug 提出了准等价(quasi-equivalence)假说。他们的想法简单地说就是：亚基在近乎相同的局部环境下与相邻亚基形成近乎等价的连接，从而使 20 面体衣壳由多个亚基进行自我装配。就高度有序的 20 面体来说，病毒粒子的对称性可以用 20 面体的三角剖分数(triangulation number)来定义(图 11-4)。三角剖分数(T)的定义为：$T=f^2\times P$。其中，f 是三角形面每一条边上的分割次数，因此 f^2 就是每个面上的小三角形数目。$P=h^2+hk+k^2$，其中 h 和 k 为任何不同的非负值整数。这就是说 T 值将以 1，3，4，7，9，12，13，16，19，25，27，28……的方式递增。当 P 等于 1 或 3 时，可以形成规则的 20 面体。任何其它的 P 值都将导致偏斜态 20 面体，即构成 20 面体的小三角形在每个面的边线上不是对称排列。目前已经测定了包括 $T=1$(微小噬菌体科，如 $\Phi X174$)，$T=3$(许多昆虫，植物及动物病毒)、$T=4$(披膜病毒)以及 $T=7$(有尾噬菌体的头部，如 λ 噬菌体)在内的各类 20 面体病毒粒子的精确结构。具有更大三角剖分数的病毒粒子，采用不同的亚基装配方式形成 20 面体的面和顶点，并含有具框架作用的内部骨架蛋白。这些因素决定了病毒粒子的装配，其中主要是通过将预先形成的蛋白结构亚基进行再次组合来完成的。

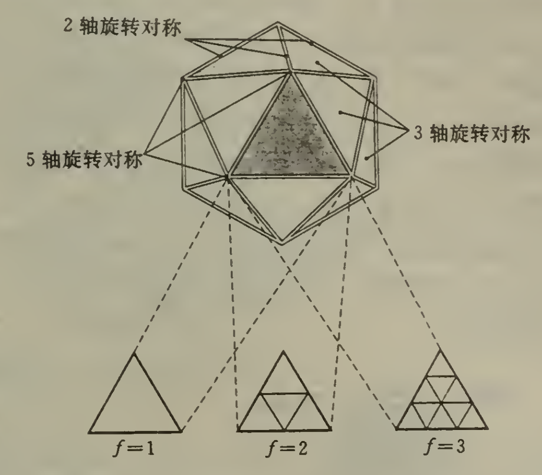


图 11-4　20 面体粒子 2-3-5 对称示意图

小 RNA 病毒(*Picornaviridae*)是 20 面体病毒粒子结构的一个很好例子。近年来，对许多不同小 RNA 病毒衣壳的细微结构进行了测定，其中包括 1 型、3 型脊髓灰质炎病毒(PV1 和 PV3)、口蹄疫病毒(FMDV)，14 型人鼻病毒(HRV-14)等许多其它种类。这些病毒的粒子结构与许多其它无关的病毒非常相似，如属于野田病毒科的昆虫病毒及豇豆花叶病毒组的植物病毒。所有这些病毒种类均为 20 面体衣壳，直径约 30 nm，三角剖分数 $T=3$ (图 11-5)。衣壳均由 60 个重复的结构亚基构成，而每个结构亚基又都含有 VP1，VP2，VP3 3 个主要的亚基。也就是说在整个小 RNA 病毒粒子中共有 $60\times3=180$ 个表面亚基单体。3 种蛋白的结构相似，都是由 150～200 个氨基酸残基以所谓的“8 股反向平行 β-折叠”构成。

实际上，小 RNA 病毒共含有 4 种结构蛋白，除上述 VP1-3 3 种主要蛋白之外，还存在第四种小蛋白 VP4。VP4 主要位于衣壳内侧，而不暴露于粒子表面。通过对小 RNA 病毒感染细胞的生物化学研究，很久以来，便已经知道这 4 种衣壳蛋白是由最初的多蛋白(polyprotein)所产生。VP4 是装配开始后由 VP0 前体切割成为 VP2 和 VP4 的，并在其氨基末端进行豆蔻酸化，即在转录后与豆蔻酸(一种 14 碳不饱和脂肪酸)发生共价结合而被修饰。5 个 VP4 单体形成一个疏水分子团，从而构成一个五聚体的结构亚基。生物化学研究结果表明，形成衣壳顶点的这些五聚体是粒子装配中的重要前体。因此可见，小 RNA 病毒衣壳组成蛋白的化学特性、结构特性以及对称性决定了其装配过程。

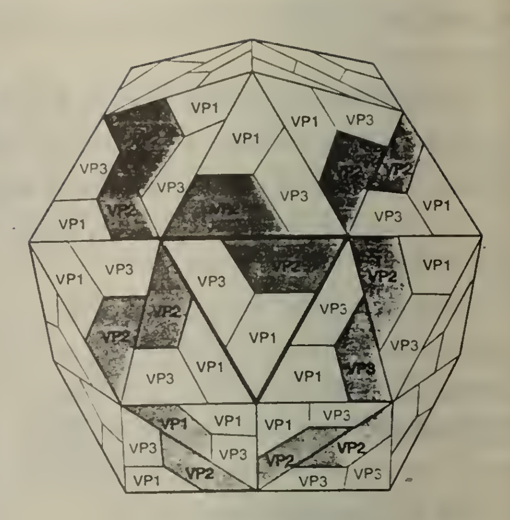


图 11-5 三角剖分数 $T=3$ 的小 RNA 病毒 20 面体粒子结构



图 11-6 包膜病毒在通过寄主细胞膜出芽时形成完整粒子

（三）包膜病毒

以上主要讲述了“裸露”病毒粒子的结构，即那些衣壳蛋白暴露于外界环境中的病毒。这些病毒在复制循环末期产生于被感染细胞，此时的细胞会发生死亡、崩溃或溶解，而将其内部组装好的病毒粒子释放出来。另外许多病毒粒子则以出芽方式(budding)穿过细胞膜，由被感染细胞内释放时仍然保持寄主细胞膜的完整性，并不造成细胞的完全毁坏。也就是病毒粒子在穿过细胞膜时被脂类膜(envelope)所包被，这些外膜产生于寄主细胞膜，成分与细胞膜相似(图 11-6)。

这种方式对病毒来说具有许多优点。包膜下层的结构可能与螺旋体或 20 面对称体结构相连，它可能在离开细胞之前或在病毒释放过程中形成。多数包膜病毒都是利用细胞膜作为装配的场所。粒子在细胞内的形成、成熟及释放多为一个连续的过程。装配场所因病毒种类而不同，除部分病毒以细胞外膜为装配场所外，许多病毒还利用如高尔基体这样的细胞质膜进行装配。另外一些在细胞核内复制的病毒(如疱疹病毒属)则可能在核膜上进行装配。以上这些病毒通常被某种形式的液泡所包被，转移通过细胞表面后再释放。

病毒粒子的类脂膜不仅具有防止水分丧失或酶类降解的功能，同时还具有生物活性和对寄主细胞上受体分子的识别功能。这些功能需要病毒合成几种类型的蛋白对其类脂膜进行修饰，并通过图 11-7 所列的 3

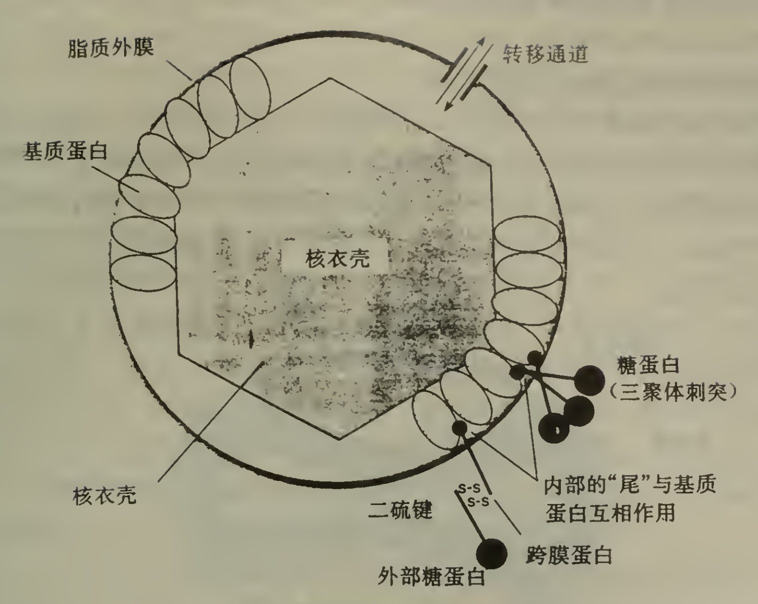


图 11-7　病毒包膜上结合的儿种蛋白质

种方式之一与包膜相连接。具体种类如下：

（1）基质蛋白是一些位于病毒粒子内部的蛋白质，它的功能是将内层的核衣壳与包膜有效的连接。这些蛋白通常不被糖基化，并经常是大量存在。如在逆转录病毒中，基质蛋白可约占到病毒粒子总重量的 30%。某些基质蛋白含有穿膜固着区，另一些基质蛋白则以其表面的疏水区域，或通过与包膜糖蛋白之间的相互作用与膜相连。

（2）糖蛋白是一些通过疏水区固着在膜上的穿膜蛋白。根据其功能又可分为以下两种类型：①借助于一个穿膜区固着在包膜上的外露糖蛋白。这种蛋白的大多数结构都在膜的外侧，仅有一个较短的尾部插入膜中。借助于电子显微技术，可在许多包膜病毒的表面上观察到由一个单体所形成的“刺突”。这些蛋白是包膜病毒的重要抗原。糖基化作用多是发生在氮原子或氧原子上的，多数这种蛋白的糖基化程度很高，总重量 75%的蛋白可能都含有转录后附加的糖基。这些蛋白不仅是包膜病毒的主要抗原，而且在病毒与外部环境发生接触中也具有许多重要的功能。例如：流感病毒的血细胞凝集素便是受体结合、膜融合以及血细胞凝集所必需的；②运输通道蛋白含有多个疏水穿膜区，可以排列构成一个穿过包膜的通道，使病毒包膜的通透性（如离子通道）发生变化。因此，这种蛋白对病毒粒子的内部环境是很重要的，它保证了病毒粒子在成熟和侵染过程，能够进行各种必需的生物化学变化，甚至是这类变化的决定因素（如流感病毒的 M2 蛋白）。

与许多具有包膜的脊椎动物病毒相比，仅有极少的植物病毒具有包膜，而且多属于弹状病毒科。除了植物弹状病毒以外，仅有少数侵染植物的布尼病毒和披膜病毒成员的外层具有类脂包膜。在植物病毒中的这种情况，可能反映了寄主细胞的生物学特性，尤其是病毒由被侵染细胞向外释放的有关机制，如植物病毒的释放要求在坚硬的细胞壁上必须要有缺口。在原核生物病毒中不存在这种障碍，因而也就有许多科的病毒具有包膜（如芽生噬菌体科及囊状噬菌体科等）。

（四）其它复杂病毒

大多数病毒都可以被归入上述 3 种结构类型之一，即螺旋对称、20 面体对称或包膜病毒。

但还有许多结构更为复杂的病毒。在这些病毒中，虽然一部分由多层蛋白及脂类构成的外壳也是遵循上面提到的基本对称规则，但它们体积较大，结构更为复杂，无法像比较简单的螺旋体或 20 面体病毒那样，直接进行数学描述。

痘病毒科便是这类病毒的一个例子。它们的粒子一般长约 200～400 nm，呈椭圆形或长方形，外层表面具有平行排列的脊状物，有时也呈螺旋状排列。这些粒子是非常复杂的，已经证明其中含有 100 种以上不同的蛋白质。在复制过程中具有两种形式的粒子，即由两层膜包被的胞外粒子和仅由一层内膜包被的胞内粒子(图 11-8)。电子显微镜下的超薄切片观察证

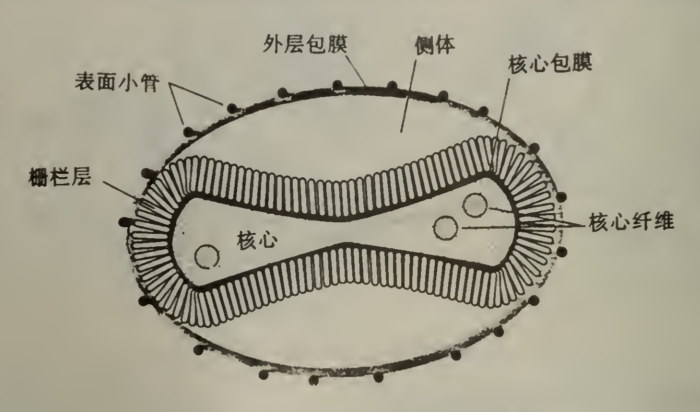


图 11-8　痘病毒的粒子结构示意图

明，病毒粒子的外层表面含有蛋白及脂类，它包被着哑铃形的髓心及两个功能尚不清楚的侧体。病毒的髓心是由紧密压缩的核蛋白构成，周围缠绕着双链 DNA 基因组。就抗原性而言，痘病毒是非常复杂的，它可以同时诱导特异性抗体和交叉反应抗体。因此，可以利用某种病毒来对另外一种疾病进行免疫接种，如使用痘苗病毒来对天花病毒进行免疫。痘病毒和许多其它病毒的复杂性还表现在它们的粒子中至少含有 10 种以上与核酸代谢或基因组复制有关的酶。除了痘病毒科以外，另外还有一些粒子结构研究较为深入的复杂病毒，如感染大肠杆菌的有鞭毛噬菌体和感染昆虫的杆状病毒。

第三节　病毒的基因组

一、病毒基因组的结构及其复杂性

病毒基因组的构成和成分的变异程度远远大于目前对整个细菌、植物或动物界研究的已知结果。构成病毒基因组的核酸可能是单链或双链；可能是一条线状或环状分子，也可能是分离存在的几个分子。单链病毒基因组可能是正意(与 mRNA 的核酸序列相同)或者负意的，也可能是双意的(上述两种的混合体)。病毒基因组的长度范围由约 3 500 个核苷酸(如轻小噬菌体科的 MS2 和 $Q\beta$ 噬菌体)至大约 280 kb(如疱疹病毒科的巨细胞病毒)。与由 DNA 构成的所有细胞生物基因组不同的是，病毒基因组含有的遗传信息仅由 DNA 或 RNA 中的一种所编码。

无论一种病毒的基因组构成如何特殊，它们都将符合上述条件之一。由于病毒是细胞内专性寄生并只能在适当的寄主细胞内进行复制，所以，病毒基因组所含的遗传信息必须以一种能够被所寄生的特定细胞体系识别并破译的形式存在。因此，病毒所使用的遗传密码必须与寄主

体系相匹配，或者至少能够被寄主所识别，而控制病毒基因组表达的指令信号也必须与寄主相适应。

由于重组 DNA 技术在病毒基因组方面的大量应用，近年来对病毒基因组的结构及核酸序列有了深入的了解。在 70 年代，第一个被测定的全基因组序列是 $\Phi X174$ 噬菌体。这种病毒具有许多特点：第一，它的基因组为已知的最小基因组之一（5 386 个核苷酸）；第二，噬菌体在大肠杆菌中可进行大量增殖，并易于纯化和抽提基因组 DNA；第三，这种噬菌体的基因组由单链 DNA 构成，可直接用于链终止法序列测定。

另外一些病毒，如疱疹病毒和痘病毒，它们含有较大的双链 DNA 基因组，复杂程度到目前为止也无法对其所有功能进行分析。许多真核生物 DNA 病毒在基因组的生物学方面与它们的寄主细胞非常相似。例如：有些 DNA 病毒基因组与细胞组蛋白结合，在病毒粒子内形成一种类似于染色质的结构。假如这些基因组存在于寄主的细胞核内，它们的作用则完全类似于小型的卫星染色体，从而对细胞内的酶类和细胞周期发生支配作用。另外，1970 年 J. Kates 在痘苗病毒 mRNA 的 3′-末端发现存在多聚腺苷酸尾序；1977 年 Sharp 在腺病毒中首次发现断裂基因存在非编码内含子、蛋白编码外显子，以及拼接而成的 mRNA。

在很大程度上，所有的病毒都受到基因组体积最小化的压力。例如，原核寄主病毒的复制速度必须达到能够与其寄主细胞的分裂保持同步的程度。许多噬菌体紧凑的结构特性也说明了这一点。基因重叠是一种普遍的现象，以这种方式，病毒得以在最小化的基因体积中包容最大量的编码信息。对于真核寄主的病毒来说，也存在对于基因组体积的选择压力，但它则主要是针对于病毒粒子的“包装体积”，也就是能够被包容在粒子内部的核酸数量。因此，与在真核细胞基因组中，由于分布密度较低而造成的遗传信息浪费相比，真核寄主病毒也通常表现出对于遗传信息的极大浓缩。

在一些例外情况中，某些噬菌体（如 T4 噬菌体所属的肌尾噬菌体科病毒）具有可达 160 kb的基因组。在这些病毒中，病毒的基因组含有许多与其自身复制有关的基因，尤其是一些与核酸代谢有关的酶。由此推断，这些病毒具有通过编码额外元件来避免受到寄主细胞生化代谢制约的能力。具备这种能力的不利一面是，它们不得不含有编码较大而复杂的病毒粒子所必需的所有遗传信息，这样也就增加了对于基因组体积的选择压力。

二、分子遗传学研究进展

一般来说，对于病毒基因组的任何研究通常都包括了以下问题。即病毒基因组的构成（DNA 或 RNA、单链或双链、线状或环状）、基因组大小及片段的数目、末端结构、核苷酸序列、编码能力（阅读框架-ORF）和调控信号（转录增强子、启动子和终止子）。

对于病毒基因组的分析可以分为两种方式：（1）基本以体外实验为主的，对于结构和核苷酸序列的物理分析；（2）对于完整病毒基因组和单个遗传因子在结构与功能关系方面的生物学研究，通常涉及病毒在活体内的表现。

对于病毒基因组的物理学分析，通常采用不同纯度的病毒提纯物，从中分离核酸。70 年代，使用逆转录酶，由 RNA 获得 cDNA 的技术被广泛采用，RNA 的序列分析得到了迅速发展。Berg 等人在 1972 年将 λ 噬菌体 DNA 的酶切片段首次克隆进入了 SV40 的 DNA 基因组中。因此，病毒基因组即是第一种克隆载体，又是第一种采用这类技术进行分析的核酸。前面已经提到，$\Phi X174$ 基因组是第一个被完整分析的复制子（replicon），此后，如 M13 等许多噬菌

体基因组也被改造用于 DNA 分析的载体。

除了分子克隆以外，其它分子水平上的分析技术也已经被大量用于病毒学研究。在采用已知标准物进行校正的情况下，利用电子显微镜技术可以直接对核酸分子的大小进行估计。凝胶电泳技术是另外一项最重要的技术。目前多采用琼脂糖凝胶来分离较大的核酸分子，而聚丙烯酰胺凝胶电泳(PAGE)则多用于较小片段的分离。凝胶电泳技术不仅可以在核酸序列分析中将大小差别仅为一个核苷酸的分子相互分离开来，而且它在完整病毒基因组的分析中也具有极高的价值，尤其是对于含有分离基因组的病毒来说。

病毒群体的生物学表现，长期以来一直被作为一种病毒学的标准技术。对于变异型病毒和自然界存在的自发突变体所进行的研究，是一种长期使用的，用于确定病毒基因功能的方法。分子生物学使得这类研究增加了一项非常有力的工具，它能够通过体外缺失和重组来设计并创造特异性的突变。尽管 mRNA 在体外体系中的转录也可以用于编码能力和某些蛋白质特性的研究，但对于病毒基因组的全部功能却只能在完整病毒上进行分析。在这方面，即便与最简单的细胞生物相比，病毒基因组的结构也更为简单，具有可以由纯化或克隆的核酸中获得侵染性病毒的特点。由核酸所引起的细胞侵染称作转染(transfection)。

当采用不含有任何蛋白的纯化病毒 RNA 对细胞进行接种时，由正意 RNA 构成的病毒基因组具有侵染性。这是因为正意 RNA 也就是 mRNA。在正常情况下，被侵染细胞内首先进行的便是病毒 RNA 翻译产生基因组复制所需的病毒蛋白。在这种情况下，向细胞中直接导入 RNA 对复制循环的早期步骤不会产生影响。由双链 DNA 所构成的病毒基因组一般也能够具有侵染性。由于双链 DNA 病毒基因组必须由寄主的复制酶转录产生 mRNA，所以涉及的机制就更为复杂一些。这种过程在感染原核生物的噬菌体基因组中较为简单，但对于像疱疹病毒那样在真核细胞的细胞核内复制的病毒来说，其 DNA 必须首先进入特定的细胞器官。虽然由转染导入细胞的 DNA 中的绝大部分都将被细胞内的核酸酶降解，但总有一小部分含有某种序列的 DNA，在进入细胞后可以很快进入细胞核，并在其中由细胞内的复制酶进行转录。另外，虽然效率低于前面提到的病毒 RNA，有些正意 RNA 病毒(如小 RNA 病毒)的基因组 cDNA 克隆也可以具有侵染性。这可能是因为细胞内的酶对 DNA 进行转录，而产生了 RNA。利用这些技术，可以由克隆的病毒基因组，包括那些在体外经过改造的基因组，重新获得病毒。

由于含有负意基因组的病毒粒子中都含有一种病毒特异性的复制酶。病毒基因组进入细胞后，负意基因组首先要由这种复制酶复制，产生像 mRNA 那样可以直接翻译的正意转录子，或者是先产生所谓的复制中间体(RI)及复制型(RF)双链分子，并以此为模板进一步合成 mRNA。由于纯化的负意基因组不能直接翻译，也不能在没有病毒复制酶的情况下复制，因此这类基因组自然情况下不具有侵染性。最近建立了一种由纯化或克隆的核酸中获得负意基因组病毒的系统。它的原理是通过向纯化病毒或病毒感染细胞的提取物中加入 mRNA，或者用 mRNA 转染“辅助细胞”(helper cell)，这些辅助细胞能够以异位(in trans)互补方式提供所需功能。这一系统的建立，使对负链病毒的遗传研究成为可能。

三、病毒的遗传学

对于病毒的遗传学研究，传统上主要是通过对于突变体的分离及分析来进行的。随着分子遗传学的建立，大量新的遗传学技术应用于不同类型的病毒。例如：使用代谢抑制剂来进行生物化学分析并构建遗传图谱，使用高分辨率电泳技术对病毒蛋白或核酸的遗传多态性进

行鉴定以及其它的分子生物学技术。通过这些分析可以建立以下不同类型的遗传图谱。

1. 基因重组(recombination)图谱　根据两个遗传标记之间的重组概率，按照一定的排列顺序进行突变，重组概率与它们之间的距离成正比。这是一种经典的遗传学技术，可用于基因组不分段的病毒(DNA 或 RNA)。

2. 基因组重配(reassortment)图谱　在分段基因组病毒中，对于特定基因组片段的突变可以用来确定与每个基因片段相对应的遗传连锁重配组合。

3. 物理图谱　转染敏感细胞后，利用小片段野生基因组对突变体基因组进行修复(如在突变体与野生型 DNA 之间形成杂种双链分子)，从而确定突变或其它变化在病毒基因组中的物理位置。

4. 限制性酶切图谱　使用限制性内切酶对 DNA 的特定位点进行切割，可以用于确定病毒基因组的结构。利用 cDNA 克隆，也可对 RNA 基因组进行分析。

5. 转录图谱　以下方法可以对编码不同 mRNA 的区域进行定位。① 使用各种 mRNA 对特定的基因组片段杂交(如限制性酶切片段)。mRNA 的确切起始或终止位置可以通过放射性探针的 S1 酶降解来确定。利用体外翻译确定各种 mRNA 的编码蛋白；② 由于紫外光造成病毒 RNA 的部分降解，在体外翻译体系中，距离翻译起始位点越远的部分被表达的可能性也就越小，因此可对 RNA 病毒基因组进行紫外辐射，来确定其开放阅读框架的位置。

6. 翻译图谱　使用翻译起始抑制剂来确定病毒的蛋白编码区域。脉冲标记使放射性仅仅掺入到那些在抑制剂加入以前起始合成的蛋白中，靠近基因组 3′端编码的蛋白标记强度最高，而靠近 5′端的编码产物则最低。

(一)病毒突变体

株系、“型”、变种、突变体以及分离物都是一些在使用中可以互换的名称，目的是将它们区别于所来源的野生型病毒。引起病毒间遗传差异的基础尚不清楚，但这些名称的准确定义如下：株系是同一病毒的不同系或分离物；型是同一病毒的不同血清型(如不同的中和反应表现型)；变种则是与原始野生型株系不同的病毒。

1. 病毒突变的类型

(1)自发突变　在有些病毒中，单个核苷酸的突变频率可高达 10^{-3}～10^{-4}(如逆转录病毒中的 HIV)，而另一些病毒的突变频率则较低，与细胞 DNA 的突变频率相近，可能只有 10^{-9}～10^{-11}(如疱疹病毒)。这种差别是由基因组的复制机制所引起的。依赖于 RNA 的 RNA 复制酶(RdRp)所产生的错误比率一般高于依赖于 DNA 的 DNA 复制酶(DdDp)。虽然有些 RNA 病毒的复制酶的确具有校对功能，但一般来讲，大多数 RNA 病毒的突变频率明显高于 DNA 病毒。就特定病毒而言，突变也是一个复杂的过程。病毒具有产生抗原性变种的能力，以避开免疫反应，无疑是有它的优点。但是大多数突变又是有害的，它也产生了许多缺损型粒子。在一些特殊的情况下(如 HIV)，单个核苷酸的错误比率可达 10^{-3}～10^{-4}。HIV 的基因组长度约为 9.7 kb，复制产生的单个基因组中将应存在 0.9～9.7 次突变。因此，所谓“野生型”病毒实际上也就是病毒复制过程中，一种在不同病毒繁殖体的动态平衡中暂时占主导地位的基因组群体。这些分子变异体混合物现在被称作“近似种”(quasispecies)，其它 RNA 病毒中(如小 RNA 病毒)也有这种情况存在。这些变异体中的大多数将不具有侵染性或有严重缺陷，因而被很快地从复制群体中淘汰出去。

(2)诱导突变　大多数病毒的遗传学分析都是用诱变剂处理后，从病毒群体中分离得到突

变体作为材料进行的。诱变剂可分为体外诱变剂和体内诱变剂两类。体外诱变剂的活性不依赖于核酸的复制。如：亚硝酸可对碱基进行脱胺，导致AT→GC及CG→TA的同型转换；羟胺可导致CG→TA同型转换；而烷化剂（如亚硝基胍）则可对嘌呤核苷的N-7位进行烷基化。

体内诱变剂的活性则针对代谢性核酸，即正在进行复制的核酸。如碱基类似物（2-氨基嘌呤、5-溴脱氧尿及5-氮胞苷），在掺入核酸的过程中，它们可导致碱基错配而发生同型转换（transitions），即嘌呤变为嘌呤和嘧啶变为嘧啶（如AT→GC和CG→TA）；或异型转换（transversions），即嘌呤变为嘧啶和嘧啶变为嘌呤（如AT→TA和GC→CG）。尿嘧啶类似物5-荧光尿嘧啶，是一种常用的RNA基因组诱变化合物。

另外，嵌和试剂（如丫啶类染料）在碱基间的交联可造成插入或缺失；紫外照射可导致嘧啶形成二聚体，并被修复功能从DNA中去除。但这种修复功能的错误比率比DNA复制过程中的多数酶类都要高得多。这一技术也适用于RNA的诱变。

诱导突变体的关键是每个基因组中发生变异的次数。所使用的诱变剂剂量必须非常谨慎，每个基因组的平均变异数目不能超过一次。否则所产生的病毒将含有多种变异，而使得对其表现型的解释更为复杂。

2. 突变的表现型

（1）噬菌斑形态　与野生型病毒相比，由于复制速度的加快或降低，突变体即可能形成较大的噬菌斑，也可能形成较小的噬菌斑。噬菌斑的大小常常与对温度的敏感性（t. s.）有关。

（2）寄主范围　这一概念在这里不仅包括了活体寄主，也包括了不同类型的体外受纳细胞（permissive cell）。在体外系统中，已经利用终止密码抑制子细胞分离获得了这种类型的条件突变体（conditional mutant），如许多噬菌体和一些动物病毒。

（3）温度敏感性（t. s.）　一类对高温表现敏感的突变体，在分离获得条件致死突变体（conditional-lethal mutant）中非常有用，也是测定有关病毒基因的有效方法。这种突变经常是由蛋白的错义（missense）所导致的。由于某些氨基酸的代换，使完整蛋白的构象产生了一些微小的变化，而只能在较低的温度下才具有功能，但突变后的蛋白质在免疫学上一般不发生变化。虽然在非适合温度下，这些突变一般仍可部分保留原有的活性，但即便在适合温度下，突变蛋白的功能也会降低。由于野生型病毒的复制比这些突变体更快，因此，这类突变经常存在着较高频率的突变恢复问题。在一些病毒中（如孤肠呼病毒和流感病毒），多年来已经获得了针对每一个基因的许多温度敏感型突变体，用于对基因组的完整遗传分析。

（4）低温敏感性（c. s.）　这种突变体与上面谈到的相反。由于寄主细胞可以在较低的温度下增殖，因此，低温敏感突变体对噬菌体和植物病毒的研究非常有用。

（5）无义突变　也称为终止突变，即在编码区内出现了三种翻译终止密码中的一种（UAG-琥珀子；UAA-赭石子；UGA-乳白子）。翻译由此提前终止，仅产生相应蛋白的氨基末端片段。这种突变体的表现型可以通过将病毒在含有不同抑制子tRNA的细胞（细菌或动物细胞）中繁殖而得到恢复。无义突变不像温度突变那样仍然可保留部分活性，仅能通过原来位点的恢复重新成为野生型病毒，但恢复频率较低。

（6）缺失突变　与无义突变近似，但可能涉及一个或多个病毒基因，也可能涉及如启动子等非编码调控区。自然缺失突变体经常以缺损型干扰（DI）颗粒的形式在病毒群体中出现。有人认为，对于某些病毒的侵染来说，这些没有侵染性，也不是病毒遗传所必需的基因组可能在侵染过程的建立及致病性方面具有重要作用。遗传缺失只能通过重组才能恢复为野生型

病毒，发生频率较低。缺失突变可以通过物理分析进行定位，所以对了解病毒基因组中结构与功能的关系非常有用。

(7)生化特征突变　其中包括了药物抗性突变；导致毒力变化的特异性突变；使蛋白或核酸电泳迁移率变化的多态性；以及对钝化剂敏感性的变化。

(8)回复突变　反向突变是突变体自身进行的一种变化，上述大多数类型的病毒都可以发生。突变形式即可以是直接地回复突变(对原有突变的修复)，也可以是其它位置上的补偿突变。

(二)病毒之间的遗传互作

寄主通常都是受到一个以上的病毒侵染，活体中的病毒之间经常会在遗传学上发生相互作用。通过培养细胞的混合侵染(superinfection)试验，遗传互作可分为以下两种类型：① 突变体按功能基团分为互补群；② 根据重组频率分析，将突变体按线性遗传图谱排列。

功能互补就是侵染细胞中病毒基因产物的相互作用。通过这种作用，虽然相关的一种或者两种突变体的基因型并未发生任何变化，但它们的复制得到了增强。也就是在混合侵染中，其中的一种病毒提供了另一种病毒所缺少的某种功能性基因产物(图 11-9)。如果两种突变体在同一功能上发生缺损，便不能在复制中相互促进，这样的两个突变体属于同一个互补群。理论上讲，互补群的数量应该与病毒基因组中的基因数目相等。但由于某些突变是致死性的，有些基因又是非必需的，无法进行测定，所以，在实际上互补群的数量总是少于基因数目。目前，已知有两种可能的互补方式，即等位(allelic)互补和非等位(non-allelic)互补。等位互补是一种基因内的互补，也就是不同突变体在同一蛋白上存在互补性缺损，如不同的功能区或多聚体蛋白的不同亚基之间。非等位互补则是基因间的互补，突变体在不同基因上发生缺损，这种形式更为常见。

重组是混合侵染中病毒基因组之间的物理作用，它使子代病毒的基因组含有与亲代病毒不同的遗传信息。根据病毒基因组的结构，有 3 种不同的重组方式。

1. 通过核酸的断裂和重新连接，产生分子内重组　以 DNA 为中间体的所有 DNA 或 RNA 病毒都会发生这种重组，并可能是在细胞中的酶作用下进行的。

2. 通过复制选择进行的分子内重组　发生在 RNA 病毒中，可能是由于在基因组合成中，病毒聚合酶在模板链上发生易位所引起的。虽然其中可能有某些细胞中的酶(如剪切酶)参与，但基本上是一种随机的过程。RNA 病毒侵染过程中的缺损型干扰粒子通常是由这种方式产生的。

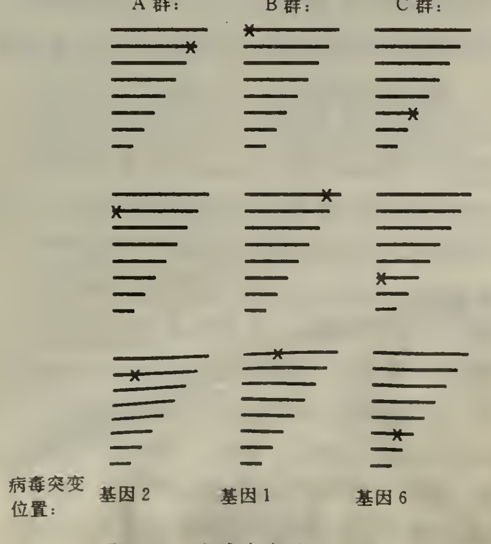


图 11-9　流感病毒的互补群

3. 重配　具有分离基因组的病毒在混合侵染中，来源于不同亲本病毒的基因组片段可能会重新组合。如流感病毒含有 8 个基因组片段，所以在混合侵染中便可能产生 2^8，即 256 种子代病毒。在这类病毒中，粒子包装过程也可能与重组子的产生有关。对于分子内重组来说，两个标记基因之间发生上述重组的可能性，与它们之间的距离成正比。所以，一对标记基因之间的重组机率最高可达 50%，并可以根据图距

单位(即重组机率百分比)的距离定位在一个线性遗传图谱上。对于重配来说，两个标记基因产生重组子的概率可能很高，也可能较低，分别代表了标记基因是位于两条不同的基因组片段上还是在同一片段上。重配作用通常会掩盖同时发生的分子内重组。

(三)病毒间的非遗传性互作

病毒之间发生的许多非遗传性互作，也可以影响遗传作用的结果。在病毒中，除了逆转录病毒整个基因组都是两个拷贝的二倍体以外，还有一些病毒(如疱疹病毒)也含有重复序列，是一种部分杂合体。另外，少量病毒(多为包膜病毒)多个基因组的异常包装，有时也可能产生杂合的多倍体粒子。如在新城疫病毒中，杂合粒子可达10%。这种杂合现象(heterozygosis)导致了病毒群体遗传的复杂性。

干扰作用(interference)是不同病毒之间另一种常见的非遗传性互作。这是在受到病毒感染的细胞中，首先侵染的病毒对后来侵入病毒的一种抵抗作用。在同种病毒之间发生的干扰作用多是由缺损型干扰粒子引起的，它们可以对必需的细胞成分形成竞争并阻断复制过程。在其它类型的突变体之间(如显性温度敏感型突变)也可以产生干扰作用。

抑制作用(suppression)是发生在病毒基因组或寄主细胞基因组中的一种抑制性突变，对另一种突变表现型的抑制。这种作用的机制不同于由寄主编码的抑制型 tRNA 引起的所谓"信息抑制"，它使遗传突变体病毒具有明显的野生型表现，也就是一种假性回复体(pseudorevertant)。这种现象在原核生物中研究最为深入。最近，在一些动物病毒中也发现这样一些例子，如孤肠呼病毒、痘苗病毒和流感病毒。由于抑制作用使病毒能够克服某些突变引起的不利影响，因此在生物学上也是非常重要的。突变体病毒可以通过以下 3 种方式之一而表现为回复：①真实回复，通过回复突变产生野生基因型和表现型；②基因内抑制，原来突变的基因经过校正后发生补偿突变，如通过另一次移码突变使原来的读码框架得到恢复；③基因外抑制，在不同的病毒或寄主基因上发生抑制突变。

在混合侵染中，子代病毒含有来自于两种亲代病毒结构蛋白的现象叫做表型混合(phenotypic mixing)。其中，有些病毒的基因组可能完全包被在另外一种病毒的衣壳或包膜内，而另一些子代病毒的外壳或包膜则可能是由两种病毒蛋白混合构成的。由于这种混合作用，虽然未使子代病毒发生任何遗传性改变，但却具有了一些依赖于粒子组成蛋白的表现型特征，如对于细胞的趋向性。这种作用在相近的无包膜病毒(如肠道病毒的不同株系)或不相关的包膜病毒之间很容易发生。例如，在有包膜的逆转录病毒中，可以通过辅助病毒，使复制缺损型病毒得到回复。也就是不同有包膜病毒的糖蛋白可以在非特异性的

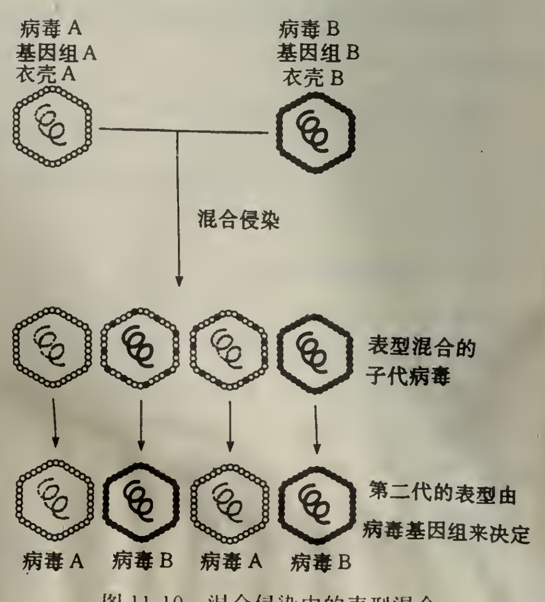


图 11-10 混合侵染中的表型混合

装配的包膜中，形成一种基因组与包膜蛋白来源不同，具有混合表型的假型病毒(pseudotype

virus)，在弹状水泡性口膜炎病毒与逆转录病毒之间也可以发生这种作用(图 11-10)。表型混合现象对于研究病毒的生物学特性是一种非常有用的工具。

四、DNA 病毒基因组

许多类型的病毒具有较大而且复杂的基因组，这些病毒在遗传中的许多方面与所侵染的寄主细胞相似。下面以腺病毒科和疱疹病毒科为例加以说明。

疱疹病毒科是一个较大的科，它含有 100 多种病毒，到目前为止在每种动物上至少发现一种。共有 7 种感染人的疱疹病毒，它们都具有基本相同的基因组，但在基因组结构和核苷酸序列上存在一些细微的差别。根据它们的生物学特性，又分为 3 个亚科，即：

α 疱疹病毒亚科：在感觉神经节表现为潜伏侵染，如人 α 疱疹病毒 1 型和人 α 疱疹病毒 3 型。基因组约 120～180 kb。

β 疱疹病毒亚科：寄主范围较窄，如巨细胞病毒和人 β 疱疹病毒 5 型。基因组约 180～280 kb。

γ 疱疹病毒亚科：感染淋巴母细胞，如人 γ 疱疹病毒 4 型。基因组约 105～170 kb。

疱疹病毒的基因组体积可达 230 kb，为双链线状 DNA，病毒粒子体积大而且较复杂，其中的蛋白质在 30 种以上。所有的病毒都编码一些不同的酶，参与核酸代谢、DNA 合成及蛋白质加工(如蛋白质激酶)。虽然这个科中的不同成员在基因组序列及蛋白质种类上差别很大，但它们在基因组的构成及结构上却是相似的(图 11-11 a)。有些疱疹病毒的基因组含有两个共价结合的部分，即分别与反向重复连接的单一长区(U_L)和单一短区(U_S)。这种重复使单一区域在结构上可以发生重排，并因此使基因组存在 4 种功能相同的异构体(图 11-11 b)。疱疹病毒的基因组中还含有一些多次重复的序列，由于重复次数的不同，同一种病毒的不同分离物的基因组差别可达 10 kb。

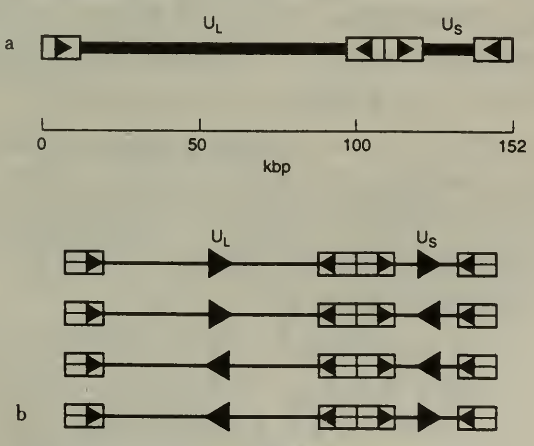


图 11-11 (a)有些疱疹病毒的基因组中含有两个共价结合的区域；(b)有可能形成 4 种不同的基因组异构形式

这个科的典型成员是人的疱疹病毒，基因组由长约 152 kb 的双链 DNA 构成，全部核苷酸都已被测定。其中至少含有 75 个基因，以非常密集的方式排列，并有读码框架重叠，每个基因都是由自身的启动子表达的。

与疱疹病毒相比，腺病毒的基因组则是由一种长度约 30～38 kb 的双链 DNA 构成，确切长度在不同种类之间有所差异，分别含有 30～40 个基因(图 11-12)。每条 DNA 的末端序列都是一种长约 100～140 bp 的反向重复，因此变性后的单链 DNA 可以形成一种“锅柄”状的结构。这种结构在 DNA 的复制中是重要的，一种 55 kD 的末端蛋白与每条链的 5′-末端共价结合，作为引物起始新的 DNA 链的合成。虽然腺病毒的基因组明显小于疱疹病毒，但是遗传信息的表达却更为复杂。不同基因共同使用少量的启动子进行表达。对产生的 mRNA 进行多次剪接，或采用不同的剪接方式，使每个启动子能够表达不同的蛋白质。

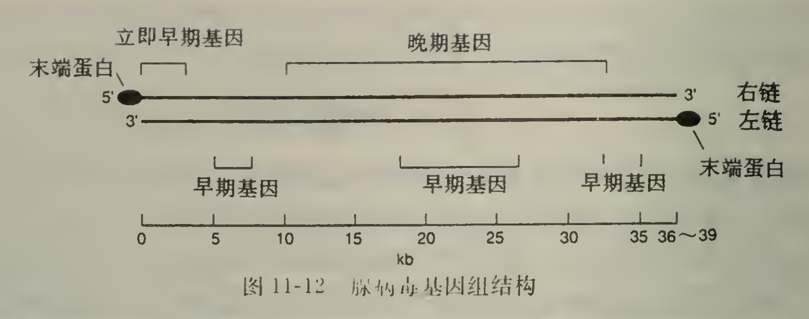
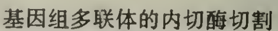


图 11-12 腺病毒基因组结构

人腺病毒被分为 6 个组(亚属)。根据分子杂交和限制性酶切图谱，病毒的基因组结构可作为将这些病毒进行归类的另一个特征。在同一组的病毒之间，全基因组的核苷酸序列同源性约 70%～95%；不同组病毒间的同源性则仅为 5%～20%。

侵染原核生物的噬菌体，在基因组结构上与其它病毒也存在许多差异，其中有些基因组小而简单；另一些却大而复杂，如 M13，T4 和 λ 噬菌体。M13 噬菌体的粒子结构与其它病毒不同，它的线状衣壳可以由于蛋白亚基的过量装配而膨胀，6.4 kb 的基因组体积也可以因额外插入的序列而增加。由于这些序列是插入在基因之间的一些不重要的区域内，所以不会影响衣壳的包装。而在其它一些噬菌体中，对于包装的要求则要苛刻得多。如在 λ 噬菌体中，只有当 DNA 为正常基因组(49 kb)大小的 95%～110%时(约 46～54 kb)才能被包装到病毒颗粒中。λ 噬菌体和 T4 噬菌体分别含有 49 kb 和 160 kb 的双链 DNA 基因组，它们代表着线状病毒基因组的另外一种普遍特征，即基因组的末端序列具有重要作用。

对于 λ 噬菌体来说，在装配过程中被包装进入噬菌体头部的是一种长的 DNA 多联体(concatemers)，它是在营养复制的后期阶段产生的。这种 DNA 明显地是被额外装进噬菌体头部的，当完整基因都被装入时，它又被一类噬菌体编码的内切酶在特定序列上切开(图 11-13)。这些酶在每条切割链的末端造成一个叫做 *cos* 的 12 bp 5′突出末端。通过这些粘性末端之间产生的氢键可以形成一种环状分子。*cos* 位点两侧的缺口可以在侵染细胞以后，经 DNA 连接酶的作用而弥合，并以这种环状 DNA 进行营养复制或整合到细菌染色体中去。



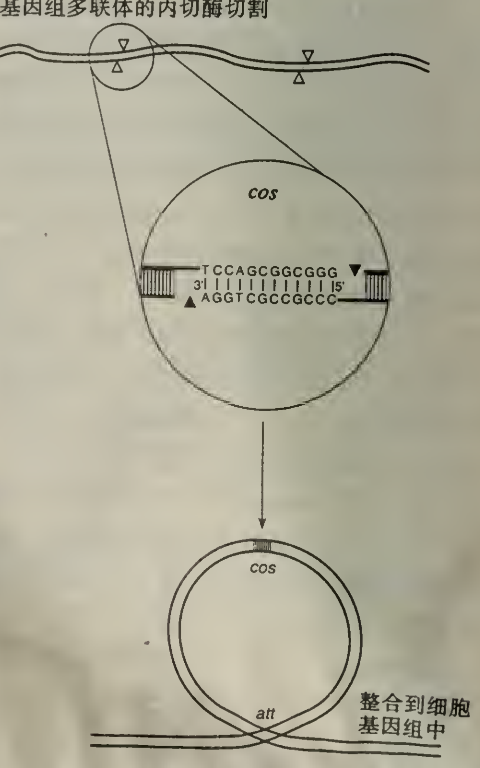


图11-13 λ 噬菌体基因组 *cos* 位点的粘着末端在寄主细胞内连接形成一种环状分子，通过对基因组中 *att* 位点的特异性识别和切割，环状分子可以整合到大肠杆菌染色体中

T4 噬菌体和其它的大肠杆菌噬菌体还具有某些线状病毒基因组的另外一种分子特征，即末端丰余(terminal redundancy)。T4 噬菌体的基因组复制也能产生长的 DNA 多联体，并被特定的内切酶所切割，但与 λ 噬菌体基因组

的不同。被装配进入颗粒的T4 DNA，在长度上稍大于完整基因组(图11-14)。也就是说基因组两个末端的某些基因发生了重复，使包装在病毒粒子内的DNA含有重复的信息。

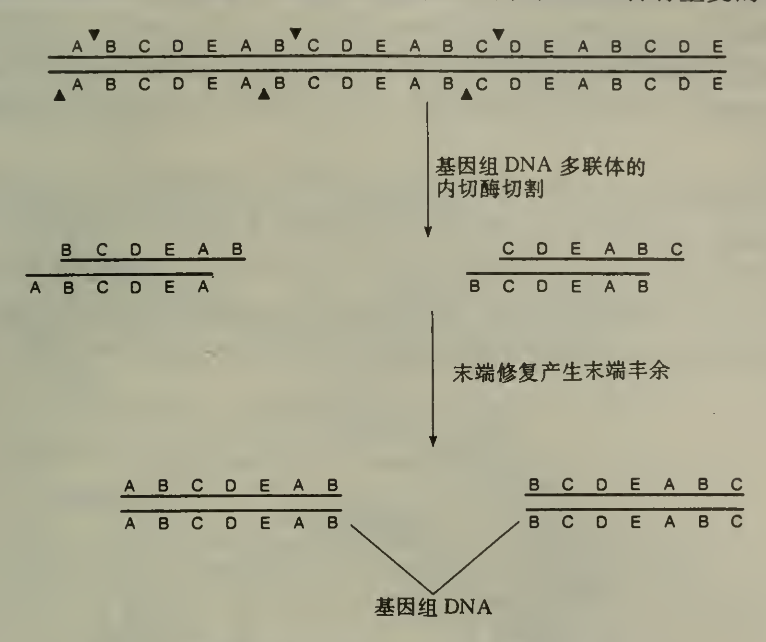


图11-14　T4噬菌体基因组末端丰余导致遗传信息的重复

细小病毒和多瘤病毒是两类动物病毒，它们代表了具有较小DNA基因组的病毒。其中细小病毒科包括三个属，即细小病毒属、依赖病毒属和浓病毒属。细小病毒的基因组为长约5 kb的线形、不分段单链DNA。所有这类病毒都至少含有一定比例的正意链，除在依赖病毒中正意链与负意链的部分相等外，其它种类包被在粒子内的核酸大部分区域为负意链。它们的基因组都非常小，即使是能复制的(replication-competent)细小病毒也只含有编码产物与转录有关的*rep*和编码外壳蛋白的*cap* 2个基因。然而，这些基因的表达却是相当复杂的，每个基因都有着类似于腺病毒那样的多种剪接方式。这类病毒的基因组末端都有约115个核苷酸的回文序列，并形成“发夹”结构(图11-15)。这种结构对于基因组的复制起始是必需的，它也体现了基因组末端序列的重要性。

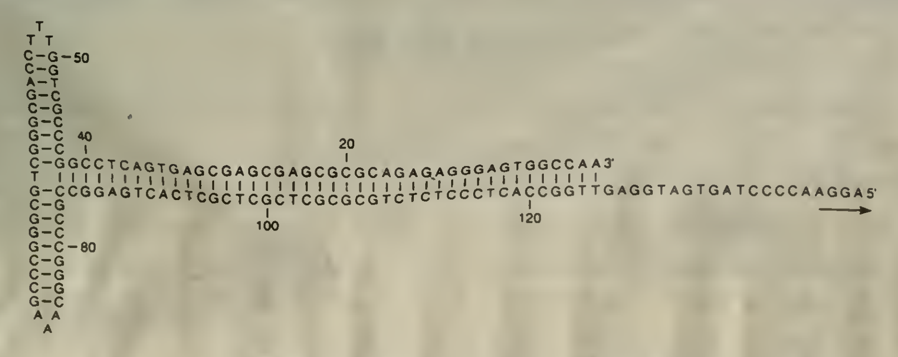


图11-15　细小病毒基因组末端的回文序列形成与复制有关的“发夹”结构

在细小病毒中，目前是根据病毒基因组的功能来划分病毒属的。依赖病毒属是首先发现

的一个属，它包括了4种类型的腺病毒伴随病毒(AAV 1～4)，也就是一些完全依赖于腺病毒(或疱疹病毒)的辅助作用才能够进行复制的病毒。虽然依赖病毒复制所必需的腺病毒基因可以编码早期调控蛋白，而不编码后期的腺病毒结构蛋白，但最近的研究结果表明，使用紫外线、放线菌酮或某些致癌剂处理细胞后，可以取代依赖病毒对于辅助病毒的需要。因此，依赖病毒所需要的可能是一种对细胞环境的修饰，而不是某种可能影响缺损型病毒基因组转录的特定病毒蛋白。

多瘤病毒的基因组由长度约5 kb的双链环状DNA分子构成，属内所有病毒的全基因组序列都已经被确定，并在分子水平上对它们的基因数目、排列方式以及调控信号和调控系统的功能等进行了非常深入的研究。在颗粒内部，病毒DNA以一种超螺旋的形式存在，并与H2A，H2B，H3和H4 4种组蛋白结合。由于同时利用了基因组的两条DNA链以及基因重叠，这些病毒以最小的空间(5 kb)包含了最大量的遗传信息(6个基因)(图11-16)。其中，VP1是

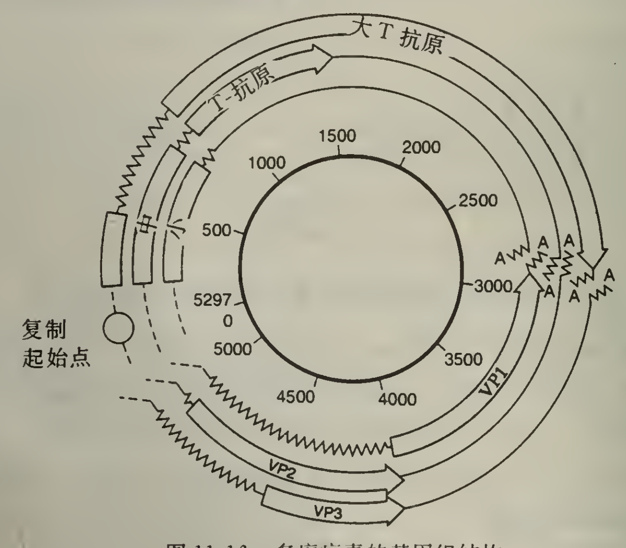


图11-16　多瘤病毒的基因组结构

由一个独立的阅读框架(ORF)编码的，但VP2和VP3基因则是重叠的，即VP3包含在VP2之中。复制起始点的两侧为控制转录的非编码区。另外，多瘤病毒还编码T抗原蛋白，这些蛋白与复制起始点结合，并表现出一些与DNA复制及病毒基因转录有关的复合活性。

五、正链RNA病毒

这一类型病毒约占已知病毒总数的70%以上，包括有包膜和无包膜、基因组分段或不分段的各种病毒。由于RNA比较脆弱，较长的核酸链更易于断裂，单链RNA基因组的大小范围因此受到限制。另外，由于复制的精确性较差，RNA基因组的突变比例高于DNA基因组。这可能是RNA病毒基因组一般都较小的另一个原因。单链RNA基因组由基因组长度约为30 kb的冠状病毒，到基因组仅约为3.5 kb的MS2和$Q\beta$噬菌体，变异幅度较大。

虽然这些病毒分别属于不同的病毒科，但它们在基因组的生物学方面都具有着一些共同的特性，即正意RNA可以直接翻译产生蛋白质，经提纯的裸露RNA对敏感寄主细胞具有侵染性，而不需要其它任何病毒蛋白的参与(侵染效率较病毒粒子低100万倍)。在基因组的复制中，这类病毒采用了与DNA半保留复制完全不同的不对称复制方式，也就是在由RNA编

码的 RNA 聚合酶(RdRp)参与下，通过双链复制中间体(replicative intermediate，RI)，产生子代基因组。由于 RdRp 对复制过程中的负意 RNA 3′-末端有着较高亲和性，复制结果不对称合成过量的正意基因组 RNA。通过对以下 4 种脊椎动物病毒科的比较，可以看出：虽然它们在基因组构成的细节上可能存在差异，但有许多相同之处(图 11-17)。

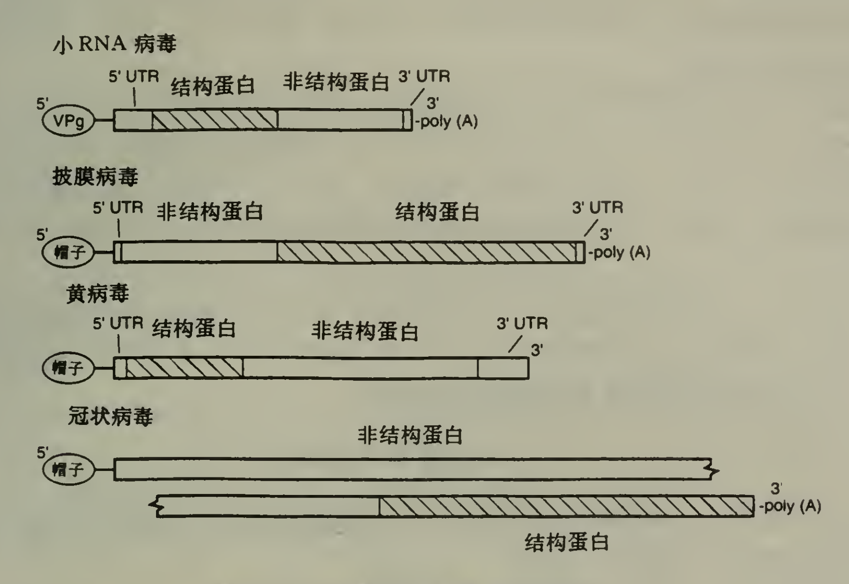


图 11-17　正链 RNA 病毒基因组结构

(一)小 RNA 病毒

小 RNA 病毒的基因组由一条单链正意 RNA 构成，长度由 7.2 kb 的人鼻病毒(HRV)到 8.5 kb 的口蹄疫病毒(FMDV)不等，并有着许多共同的特性。

(1)5′-末端含有一个长的非翻译区(UTR，600～1 200 个核苷酸)，这一区域对于病毒的翻译、致病性及衣壳装配都具有重要作用。另外还有一个较短的 3′-末端非翻译区(50～100 个核苷酸)，它是复制过程中的负意链合成所必需的。

(2)基因组的其余部分仅编码一个氨基酸数目在 2 100～2 400 之间的多蛋白(polyprotein)。

(3)基因组的 2 个末端都会受到修饰，其中 5′-末端共价结合着一个由 23 个氨基酸构成的 VPg(viral genome associated protein)碱性蛋白；3′-末端则含有多聚腺苷酸[poly(A)]。多聚腺苷酸序列并不是自身编码的，而是在 3′-末端上游有一个多聚腺苷酸信号，就像真核生物的 mRNA 那样。

(二)披膜病毒

披膜病毒的基因组由长度约为 11.7 kb，不分段的单链正意 RNA 构成，它们具有以下特征：

(1)含有与细胞 mRNA 相似的 5′-甲基化“帽子”结构和 3′-poly(A)尾序。

(2)通过两次翻译进行表达，首先由基因组 5′-区域编码产生非结构蛋白，然后再由复制而来的 3′-区域亚基因组(subgenomic) mRNA，编码结构蛋白。

(三)黄病毒

黄病毒基因组含有一条单链正意 RNA 分子，长约 11.5 kb。它们具有以下特征：

(1)5′-末端有甲基化的“帽子”结构，但 RNA 3′-末端没有 poly(A)。

(2)基因组结构与披膜病毒不同，结构蛋白由基因组5′-区域编码，而非结构蛋白则由3′-区域编码。

(3)基因表达与小RNA病毒相似，也是产生一种多蛋白。

(四)冠状病毒

冠状病毒基因组由不分段的单链正意RNA构成，长度约27～30 kb，是最长的RNA病毒。它们具有以下特征：

(1)具有5′-甲基化"帽子"和3′-poly(A)结构，病毒RNA可以作为mRNA直接表达。

(2)基因组5′端约20 kb的片段首先翻译产生一种病毒的聚合酶，再由这种聚合酶产生全长的负意链。利用这种全长的负意RNA为模板，转录产生一系列相互重叠的mRNA，在它们的5′端都具有一个由72个核苷酸构成的相同非翻译前导序列，3′端也都有一个poly(A)尾序。

(3)每种mRNA都是单顺反子，位于基因组5′端的基因由最长的mRNA进行翻译，并以此类推产生多种蛋白。这些病毒所特有的、类似于细胞质mRNA的结构不是由剪接(转录后修饰)所产生的，而是由特定的聚合酶通过转录而来。

六、负链RNA病毒

与上面讨论的正链病毒相比，具有负意RNA基因组的病毒之间差异更大一些。可能是由于表达难度的原因，这些病毒一般都有着较大的基因组，以编码更多的遗传信息。同时，在这类病毒中普遍存在着分段基因组(图11-18)。经过提纯的这类病毒基因组RNA都不具有侵染性。这是由于在未受感染的细胞中，不含有RNA编码的RNA聚合酶(RdRp)，负意基因组也不能作为mRNA进行翻译，所以这些基因组就不具有生物活性。下面要讨论的一部分病毒，并非严格意义上的负意病毒，而是一种双意(ambisense)基因组，也就是部分为负意，部分为正意。

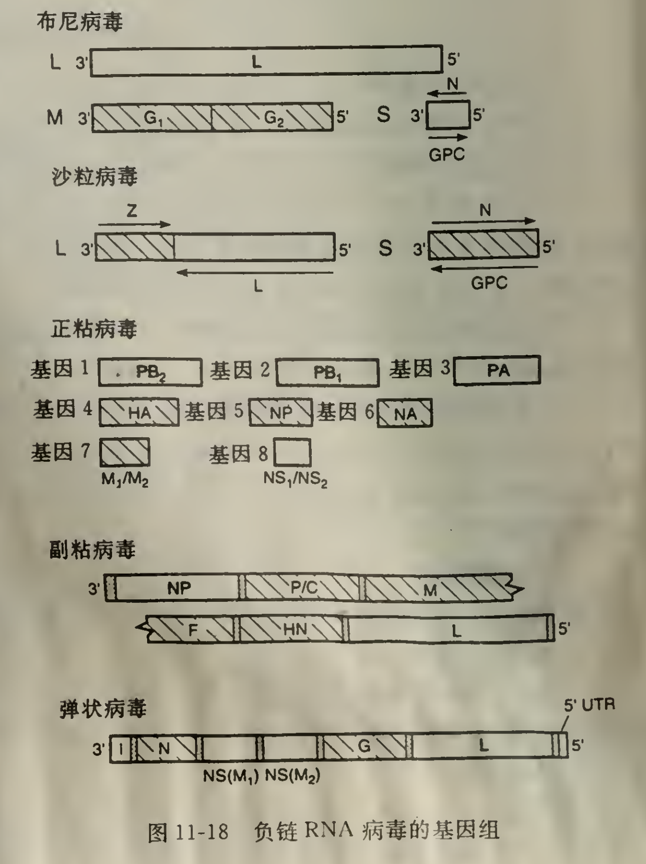


图11-18 负链RNA病毒的基因组

(一)布尼病毒

布尼病毒科成员的基因组为单链负意分段RNA，它们具有以下特征：①基因组由长(L，8.5 kb)、中(M，5.7 kb)、短(S，0.9 kb)3种分子构成;②所有的3种RNA均为线状。在病毒颗粒内，由于碱基配对使它们的末端相互靠拢而呈环状。3种RNA组

分的分子数目在病毒提纯物中不等量存在；③像所有的负意 RNA 一样，基因组 5′和 3′-末端不存在“帽子”结构或 poly(A)尾序；④白蛉热病毒属成员与本科中 S 基因组分长度较短的其它 4 个属(布尼病毒属、内罗病毒属、汉坦病毒属及托斯波病毒属)不同，它的各个基因组分都是 5′端为负意、3′端为正意的双意分子。

(二)沙粒病毒

沙粒病毒为线状单链 RNA，基因组双意，与布尼病毒科的白蛉热病毒属相似。基因组分为两段，分别为长(L，5.7 kb)和短(S，2.8 kb)组分。

(三)正粘病毒 (见分段基因组)

(四)副粘病毒

副粘病毒科成员的基因组为长度约 17～20 kb 的不分段负意 RNA。所含的 6 个基因按顺序排列，相互之间由基因末端的 poly(A)加尾信号、基因间序列 GAA 和下一个基因的翻译起始信号所构成的重复区分隔开来。

(五)弹状病毒

弹状病毒科成员的基因组为长度约 11 kb 的不分段负意 RNA。病毒 RNA 基因组的 3′和 5′-末端分别有一个由约 50 个核苷酸构成的前导区和一个由 60 个核苷酸构成的非翻译区(UTR)。所含 5 个基因的排列方式基本上与副粘病毒相似，每个基因的末端都有一个保守的 poly(A)信号，基因之间也都有一个短的间隔区。

七、分段和多组分病毒基因组

这两种类型的基因组结构经常被混淆。分段(segmented)病毒基因组是基因组被分为两个或两个以上不同的核酸分子，但被包装在同一个病毒粒子内；而多组分(multipartite)基因组虽然也是分段的，但它们的每个基因组片段却是包装在不同的病毒颗粒中的。虽然多组分病毒的不同颗粒在结构上是相似的，并可能由相同的蛋白质构成，但由于所包含的病毒基因组片段的长度不同，经常在颗粒的大小上存在差异。采用分段基因组策略的病毒很多，在人、动物和植物的病原物中都有存在，如正粘病毒、孤肠呼病毒和布尼病毒。而多组分基因组病毒则要少得多，并全部都是植物病毒。其中包括含有两个基因组分段并有两种病毒粒子的二分体(bipatite)病毒，如单链 DNA 的双联病毒组、双链 DNA 的隐性病毒组和单链 RNA 的豇豆花叶病毒组、烟草脆裂病毒组、大麦黄色花叶病毒组及真菌传杆状病毒组。含有 3 个基因组片段和 3 种病毒粒子的三分体(tripartite)病毒，如单链 RNA 的雀麦花叶病毒组、黄瓜花叶病毒组、烟草线条病毒组和大麦病毒组。另外还有一些含有更多组分的植物多组分病毒，如含有 5 个单链 RNA 基因组片段的甜菜坏死黄脉病毒。

病毒基因组的分离存在既有优点也有一些缺点。由于核酸的物理特性，尤其是较长的分子容易被剪切断裂，因而不分段病毒基因组的大小受到限制，使被包装到衣壳中的核酸不能过长。因为与双链 DNA 相比在化学性状上更不稳定，核酸链的断裂对单链 RNA 来说尤其突出。最长的单链 RNA 基因组是长约 30 kb 的冠状病毒，而最长的双链 DNA 病毒基因组则要长得多，如约 280 kb 的巨细胞病毒。由于不能进行完整的转录、翻译及复制，基因组的断裂会使病毒丧失生物学活性。基因组的分离存在使病毒避免了将其全部遗传信息过分集中，降低了由于剪切而造成断裂的可能性，从而使整个基因组所具有的编码能力得到提高。这种策略的缺点在于所有的单个基因组片段必须被包装到同一个病毒颗粒中，否则便会由于遗传信息

的丢失而造成病毒缺损。对这种包装如何进行调控的机制目前尚不清楚。

将基因组分段分别包装到不同病毒颗粒中的多组分策略，虽然不需要对每个基因组片段的包装进行精确地控制，但在有效侵染的建立中，单个寄主细胞必须受到所有不同的病毒粒子的感染。由于在植物病毒进行感染的许多途径中，都可以使大量的侵染性病毒颗粒得以侵入，从而使最初的细胞有机会受到一个以上病毒粒子的侵染，如利用获毒昆虫或对组织造成物理损伤进行的接种。这可能也是为什么多组分病毒只发生在植物上的原因。

除了基因片段的重配以外，分段基因组的遗传学与不分段基因组基本相同。无论各种基因片段是包装在一个粒子还是包装在多个粒子中，都可能发生重配。重配作用是快速实现遗传分化的有效途径，也可能是进化的另一个因素。同时基因组的分离也必然涉及到遗传信息的分割及其表达方式。

要了解这类基因组的复杂性，就必须认识分段病毒基因组(如正粘病毒科的流感病毒)和多组分基因组(如双联病毒)的结构。流感病毒 A,B 株系的基因组由 8 个单链负意 RNA 片段构成(图 11-18)，C 株系则由 7 个片段构成。所有 8 个片段的 5′和 3′-末端都具有基因组复制所必需的相同核酸序列(图 11-19)。这些序列之间是互补的。在病毒颗粒内，基因组片段的末端以碱基配对结合到一起，形成一种被认为是与复制有关的“锅柄”状结构。这些 RNA 基因组片段并非是以裸露核酸的形式被包装的，而是与第五个片段的编码产物——核蛋白相结合，在电子显微镜下形成一种螺旋状结构。就生物化学和遗传学来说，每个基因组片段都是独立起作用的，但在电子显微镜下，经非离子去垢剂解离后的流感病毒颗粒的核衣壳却仅是一种长的螺旋体结构。这说明在这些基因组片段之间存在某种相互作用，它使流感病毒的每个粒子能够以非常低的错误比率，对基因组片段进行选择和包装。决定这种精确过程的基因组序列和蛋白质相互作用尚不清楚。

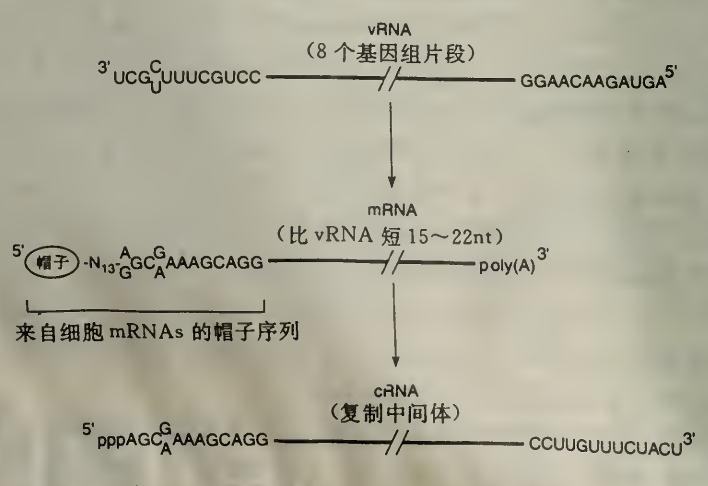


图 11-19 流感病毒 RNA 组分的共有末端序列

根据寄主植物(单子叶或双子叶)和昆虫介体(叶蝉或飞虱)的不同，双联病毒被分成 3 个组。在 I,II 两个组中，基因组由长约 2.7 kb 的单链 DNA 构成。虽然在感染细胞内的正意和负意链都含有蛋白质编码序列，但一般还是认为包含在病毒粒子内的是正意 DNA。III 组双联病毒是二分体的，含有两个单链环状 DNA 分子，分别包装在两个不同的粒子中(图 11-20)。构成基因组的两条链都是约长 2.7 kb，除了与 DNA 复制有关的一段长约 200 个核苷酸

的区域以外，相互间的核酸序列完全不同。两条基因组DNA包装在完全分离的衣壳中。由于有效侵染的建立同时需要基因组的两个组分参与，所以必须至少要有两个分别含有一个基因组片段的病毒颗粒才能对寄主细胞造成侵染。虽然双联病毒并不在它们的昆虫介体内增殖(非增殖型传播)，但摄毒量却相当大，在注入新的寄主植物时，可以使两种粒子同时造成感染。

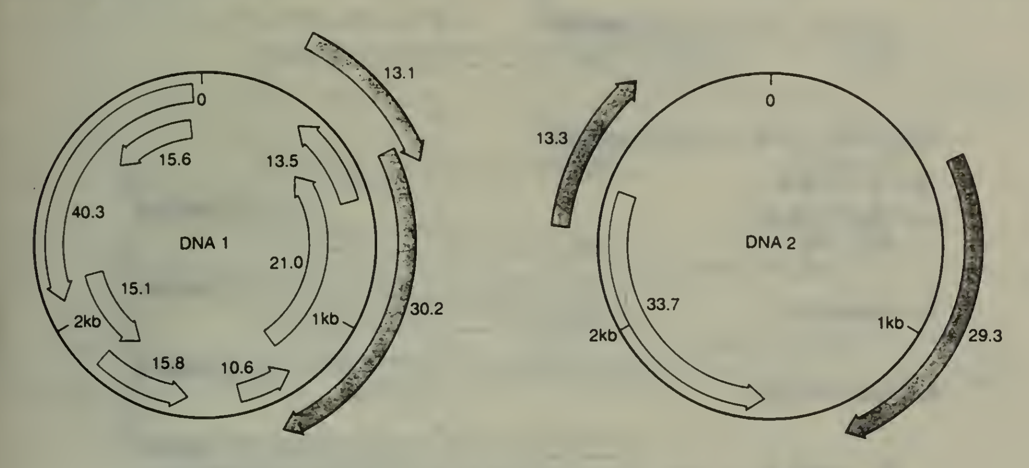


图 11-20　二分体双联病毒基因组的结构和可能的蛋白编码区

顺时针箭头为正意 ORF；逆时针箭头为负意 ORF。

八、逆转录病毒

1970 年，Howard Temin 和 David Baltimore 同时发现了在含有 RNA 基因组的逆转录病毒的粒子中含有一种依赖于 RNA 的 DNA 聚合酶(RdDp)，也就是逆转录酶。通过对逆转录病毒基因组及其复制过程的研究，目前已经获得了大量关于逆转录转座子(retrotransposon)的知识。在高等生物的基因组中，也存在着非常相似的结构。

在逆转录病毒中最引人注意的便是人类免疫缺损性病毒(HIV)，俗称艾滋病毒(AIDS)。它与其它逆转录病毒不同，除了通常的 *gag*，*pol* 和 *env* 基因以外，基因组中还存在 5～6 个其它的基因(图 11-21)。这些基因都比较小，分别通过一组经过剪切的 mRNA 表达，从而使 HIV 对被感染细胞比其它逆转录病毒具有更强的损伤。不同 *HIV* 基因的确切功能尚不清楚，但关于它们在转录及 mRNA 剪切中的作用，已经获得了一些较为深入的研究结果(见本章第五节)。

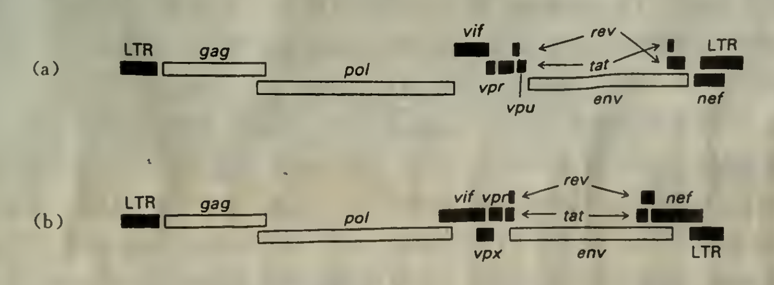


图 11-21　两种不同的艾滋病毒[(a)为 *HIV*-1；(b)为 *HIV*-2]的基因组结构

逆转录病毒是一种真正的二倍体病毒。其粒子中除了含有两个基因组RNA 拷贝以外，还含有一种在基因组逆转录合成DNA 过程中作为引物的细胞 tRNA(图 11-22)。

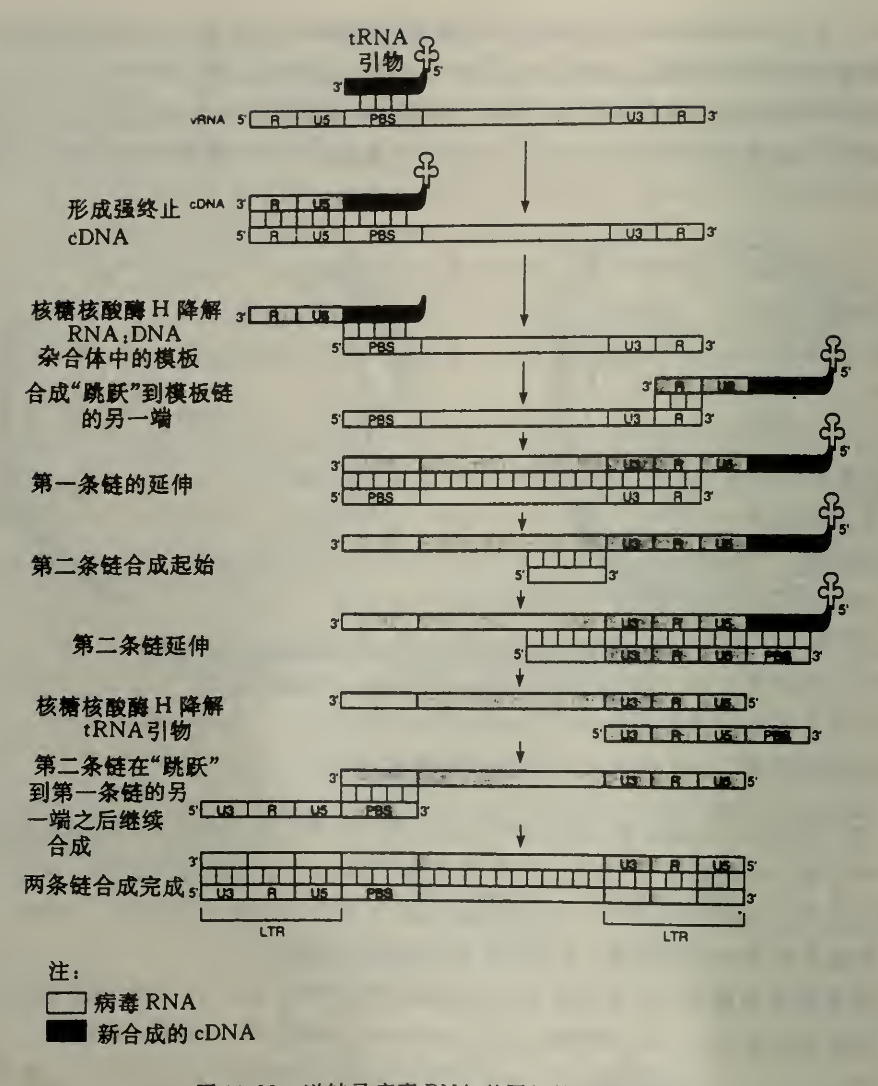


图 11-22　逆转录病毒 RNA 基因组的逆转录过程

在逆转录过程中，构成病毒基因组的两条单链正意 RNA 被转换成为一条双链 DNA 分子。由于两端正向重复序列(LTR)的延伸，所产生的 DNA 分子比原来的 RNA 模板稍长(图 11-23)。逆转录过程中的某些步骤尚不清楚，如逆转录酶由模板链的一端向另一端的长距离跳跃。由于逆转录病毒的 RNA 只能在部分脱壳的病毒核心颗粒中完全转换为双链 DNA，而在体外溶液中却不能复制，说明逆转录病毒核衣壳内的两条 RNA 的构象决定了逆转录过程。由于在核心颗粒内部病毒核酸链的末端可能是相邻的，所以实际上可能也就不存在所谓的“跳跃”。

逆转录对于逆转录病毒的遗传具有重要的作用。首先，由于逆转录酶不像细胞中依赖于 DNA 的 DNA 聚合酶(DdDp)那样具有校正功能，所以它是一种高度的易错过程。在这一过程中，逆转录病毒的基因组中产生许多突变，并导致快速的遗传分化。另外，逆转录对遗传重组也有促进作用。由于在每个病毒粒子中包装的两条 RNA 都被用作逆转录模板，所以在这两条链之间便可以发生重组。由于突变可以产生两条不同的 RNA，重组后将产生一种在遗传学上不同于任一亲代的病毒。

当逆转录完成以后，双链 DNA 进入细胞核中，并仍与病毒蛋白相结合。*pol* 基因的成熟产物实际上是一种多肽复合体，它包括了 3 种不同的酶活性，即与逆转录有关的逆转录酶和核糖核酸酶 H（RNase H），以及与催化病毒 DNA 与寄主细胞染色质整合的整合酶。整合以后才能产生原病毒（provirus）（图 11-24）。逆转录之后，在感染细胞中可以发现 3 种形式的双链 DNA，即一种线状 DNA 和两种含有一个或两个 LTR 结构的环状 DNA。根据原病毒的末端结构，曾经认为这种由两个 LTR 构成的环是用于整合的结构。近年来，

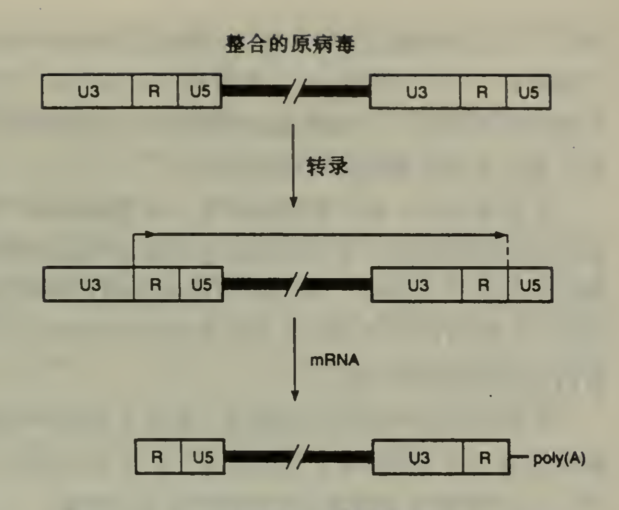


图 11-23　逆转录病毒末端重复序列的产生过程

在 U3 和 R 区域内分别含有转录启动子和多聚腺苷酸信号。

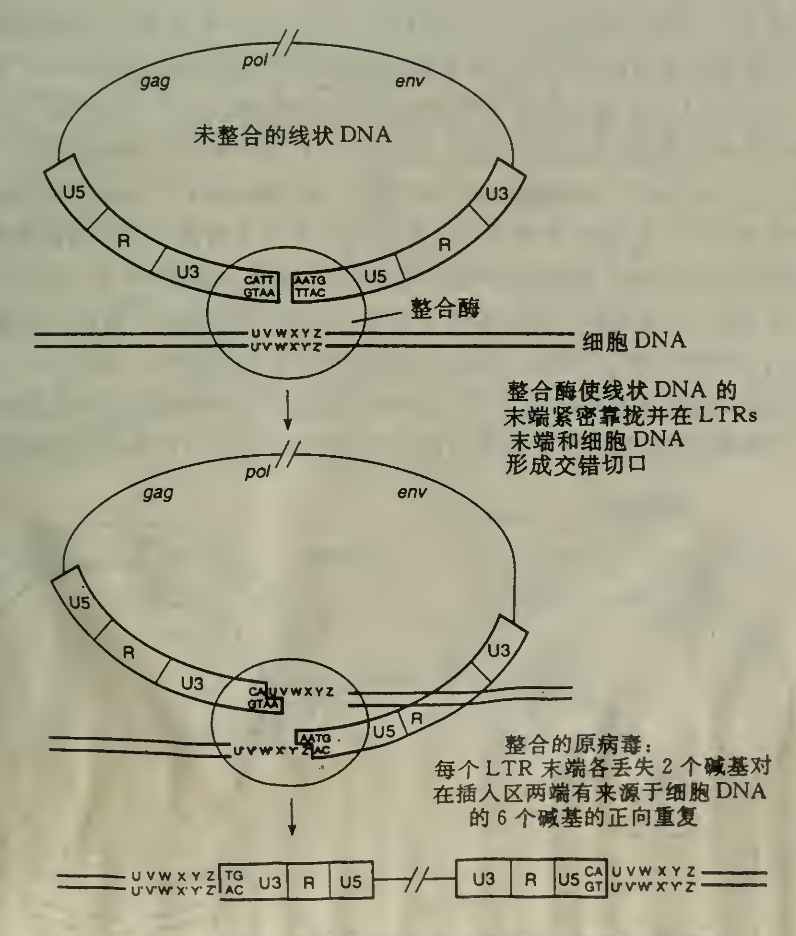


图 11-24　逆转录病毒基因组与寄主细胞染色质的整合机制

利用逆转录病毒 DNA 体外整合体系进行的研究表明，发生整合的实际上是一种线状结构。即

两个 LTR 的末端在逆转录酶-整合酶复合体的作用下，相互非常靠近。整合作用使每个 LTR 末端缺失 1～2 个碱基，并在原病毒的两端增加了来源于细胞 DNA 的 4～6 个碱基。目前，尚不清楚是否存在对于细胞基因组整合位点的特异性，但在真核生物基因组中可能存在一些比其它部分更利于整合的区域或位点。

在整合以后，原病毒基因组成为细胞基因的一部分，并随细胞的表达而表达。由于细胞基因组的修饰作用，有时原病毒也可能丢失或发生变化，关于整合后的原病毒被细胞基因组删除的机制尚不清楚。病毒由细胞基因组中释放的唯一途径就是通过转录产生一种全长的 mRNA（在 LTR 中不含有末端丰余序列）。这种 RNA 便是病毒 RNA，每个病毒粒子中都包装有两个这样的拷贝（图 11-23）。

各种具有逆转录过程的病毒，采用了两种不同的基因组策略。一种是上面提到的逆转录病毒所采用的策略，整个过程以 RNA 作为病毒基因组被包装进病毒粒子为结束；另一种是 DNA 肝炎病毒和花椰菜花叶病毒所采用的策略。它与逆转录病毒复制过程正好相反，首先进行 DNA 转录，然后在病毒粒子形成过程中再进行 RNA 的逆转录，最终产生 DNA 病毒基因组。

乙肝病毒（HBV）是 DNA 肝炎病毒科的典型代表，具有含有脂类的球形粒子，直径为 42～47 nm。基因组为有缺口的（gapped）双链 DNA，并在粒子中含有一种逆转录酶（图 11-25 A）。DNA 肝炎病毒的基因组在所有已知病毒中是最小的，负意链长度因 DNA 肝炎病毒的种类不同约为 3.0～3.3 kb，正意链长度则因病毒粒子的不同约为 1.7～2.8 kb。在对细胞的侵染中，可以产生 3 种主要的基因组转录物，分别为 3.4，2.4 和 2.1 kb 的 mRNA。所有这些 RNA 都是以同一方向，由同一条核酸链转录而来的，具有共同的 3′末端，但 5′末端不同。虽然这些转录物的长度不等，每种转录物编码那些蛋白质也尚不清楚，但已知在病毒中有以下 4 个基因，即编码核心蛋白的 *C* 基因；编码逆转录酶的 *P* 基因；编码 Pre-S1，Pre-S2 和 S 3 种表面抗原蛋白的 *S* 基因；以及编码一种病毒转录所需反式激活蛋白的 *X* 基因。侵染后不久便可在寄主细胞核内发现闭合环状 DNA，可能由它再转录产生上述 RNA。这种 DNA 是按照以下方式，通过对有缺口病毒基因组的修复而产生的：①正意链延伸合拢；②由负意链和正意链中分别除去蛋白及寡核苷酸引物；③切除负意链末端的多余序列；④两条链的末端进行连接。在

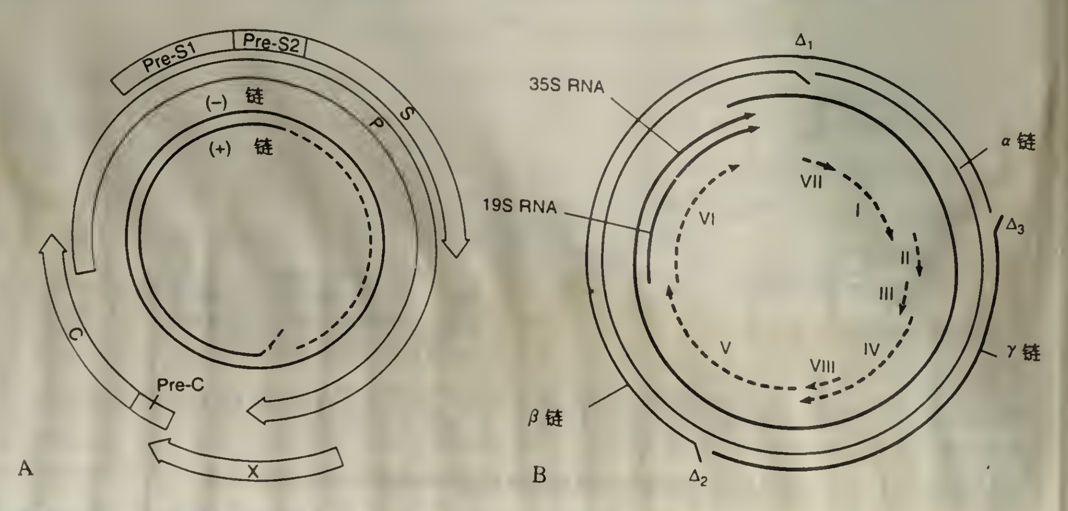


图 11-25 乙肝病毒和花椰菜花叶病毒的基因组结构和可能编码的蛋白质（A 为 HBV；B 为 CaMV）

病毒的核心颗粒中含有 3.4 kb 的 RNA、核心抗原和聚合酶，当这些颗粒在细胞质中形成时，RNA 由聚合酶转化成为 DNA。

花椰菜花叶病毒(CaMV)是花椰菜花叶病毒组的典型成员，与 DNA 肝炎病毒相比，虽然存在一些差异，但它们在基因组结构和复制中都存在相似之处。CaMV 的基因组由一条长约 8 kb，具有缺口的双链 DNA 分子构成，其中包括只含有一个缺口的 α 链和一条有两个缺口的互补链(图 11-25 B)。在它的基因组中共有 8 个基因，但尚未在感染细胞内检测到全部这些基因的产物。CaMV 基因组的复制与 HBV 相似，带有缺口的病毒 DNA 也是要首先进入感染细胞的核内，再修复成为一种共价闭合环状分子。这种 DNA 转录产生长短两种带有多聚腺苷酸尾序的 RNA。较短的 mRNA 在细胞质中翻译产生一种蛋白质，这种蛋白在侵染细胞的细胞质中形成较大的内含体(inclusion body)，它也是第二个复制阶段的进行场所。在这些复制复合体中，较长的 mRNA 进行翻译和逆转录，并被包装到正在形成的病毒粒子中。

第四节　病毒的复制

病毒的复制过程可以被分为 8 个阶段(图 11-26)。虽然这些"阶段"纯粹是人为的划分，以用来说明一种并不存在的典型的病毒复制过程，但所有的病毒，无论它们的寄主是什么，在完成其复制循环的过程中，都一定以某种方式经历了这些阶段。以下所描述的各个阶段，并非在每一种病毒中都是可以被分别检测到的。它们经常互相交织在一起，或者几乎是同时发生。其中某

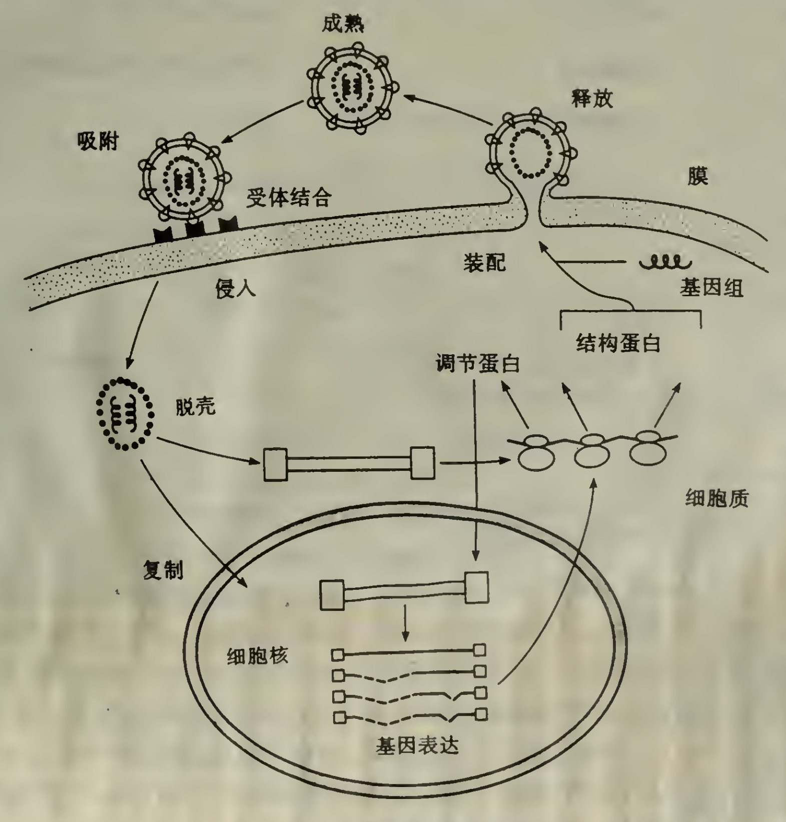


图 11-26　病毒复制过程示意图

些阶段已经被研究得非常深入，并获得了大量的资料，而另一些阶段则所知甚少或一无所知。

一、吸　　附

具体来讲，病毒的吸附是病毒附着蛋白（virus-attachment protein）（或称作反受体-antireceptor）与细胞受体（receptor）分子的特异性结合过程。现在已经发现了许多受体蛋白的例子（表 11-1）。细胞表面的受体分子可以是蛋白（多为糖蛋白），也可以是糖蛋白或糖脂中的一些碳水化合物残基。糖蛋白多为特异性受体，既病毒以一种特定蛋白为受体。由于许多不同糖基化的膜结合分子可能具有相同的侧链构型，碳水化合物基团的特异性较低。有些复杂的病毒（如痘病毒和疱疹病毒）可能采用一种以上的受体，并通过不同的途径进入细胞。

表 11-1　目前已知的一些病毒受体及其相应的病毒附着蛋白

病　　毒	受　　体	病毒附着蛋白
RNA 病毒		
小 RNA 病毒		
脊髓灰质炎病毒	类似于免疫球蛋白的 100 kD 蛋白家族	VP1
人类鼻病毒	胞内固着分子 1（ICAM-1）	V1，2 和 3（在病毒衣壳表面形成凹区）
呼肠孤病毒		
呼肠孤病毒 3	肾上腺素受体或唾液糖蛋白	σ-1 蛋白 C 末端区域（血细胞凝集素）
披膜病毒		
西门利启森林甲病毒	Ⅰ型 HLA 或 H-2 MHC（?）	E2 包膜糖蛋白
正粘病毒		
流感病毒	唾液酸寡糖	血细胞凝集素（HA）蛋白
副粘病毒		
仙台病毒	唾液酸寡糖	血细胞凝集素-神经胺酸酶（HN）蛋白
弹状病毒		
狂犬病毒	乙酰胆碱受体或唾液酸神经节苷脂	病毒 G 蛋白
逆转录病毒		
HIV-1 和 HIV-2	CD4 分子（及辅助蛋白）	gp120 包膜糖蛋白
维斯那病毒	Ⅱ型 MHC（?）	包膜糖蛋白
DNA 病毒		
腺病毒		
人类腺病毒	Ⅰ型 HLA MHC 分子	纤丝蛋白
疱疹病毒		
巨细胞病毒	Ⅰ型 HLA MHC 分子及 β_2-小球蛋白	H301 蛋白的 α 区（病毒编码的类Ⅰ型 MHC 分子）
EB 疱疹病毒	C3d 受体‘CR2’（=CD21）	gp350
痘病毒		
痘苗病毒	表皮生长因子受体（EGF）	VGF 蛋白
DNA 肝炎病毒		
乙肝病毒	多聚血清白蛋白、多聚 IgA 或唾液糖蛋白肝细胞受体	包膜蛋白的原 S 区

小 RNA 病毒科和流感病毒属是目前对于病毒与受体相互作用研究最为深入的两类病毒。小 RNA 病毒与受体的相互作用已经从两个方面得到了验证，即与受体结合有关的病毒结构特征及受体本身的结构特征。人类鼻病毒受体分子 ICAM-1 是一种细胞附着分子，它的正常功能是将细胞与附近的底物结合到一起。ICAM-1 在结构上与免疫球蛋白分子相似，也含有与抗体类似的保守区（C）和易变区（V），被归属于免疫球蛋白家族成员（图 11-27）。脊髓灰质炎病毒的受体与它相似，是同属这一家族的一种膜整合蛋白，含有一个易变区和两个保守区。脊髓灰质炎病毒受体的正常功能尚不清楚。

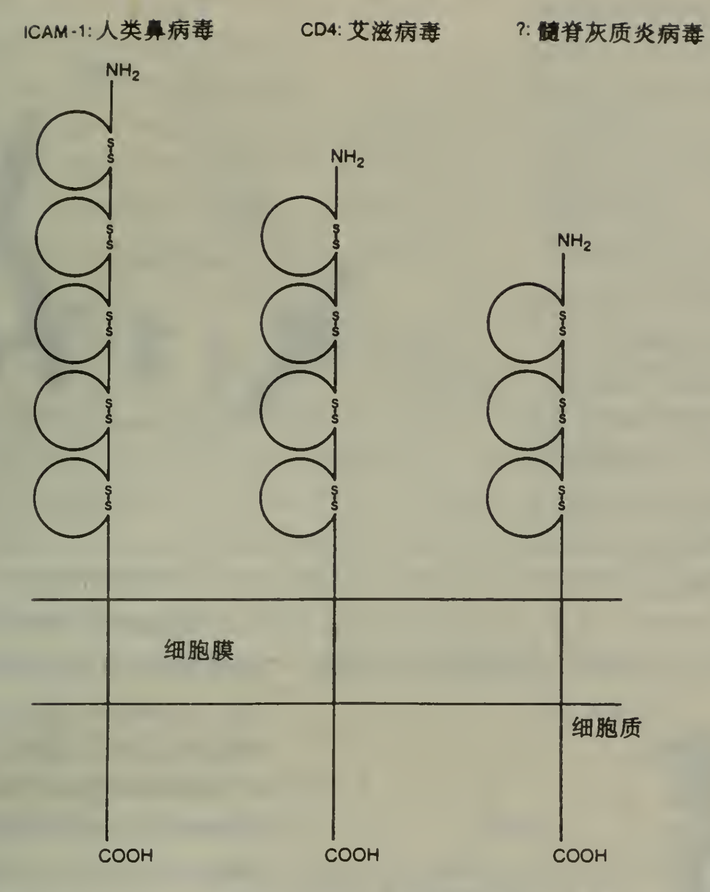


图 11-27　类似免疫球蛋白的病毒受体结构示意图

对许多小 RNA 病毒衣壳结构的认识可以达到零点几个纳米的分辨率，因此，有可能确定病毒中与受体结合相关的特征。在鼻病毒中，20 面体衣壳的每个三角形表面都存在一种被称为“凹谷”的深裂，它是由 VP1，VP2 和 VP3 3 种亚基蛋白单体之间的空隙所形成的（图 11-28，图 11-29）。利用可以阻断鼻病毒与细胞吸附的抑制剂所得到的研究结果表明，ICAM-1 与病毒粒子间的相互作用便是在这种受到遮蔽的峡谷底部进行的。与病毒表面的其它部分不同的是，构成峡谷内侧表面的氨基酸残基不易发生变异。由于抗体分子较大而不能插入这种狭缝，因此这些残基可以免受抗原性压力。虽然这一区域的基团通过变异也可使病毒回避免疫应答，但同时也将阻断与受体的结合，因而这种结构特点是非常重要的。在脊髓灰质炎病毒中，取代“狭谷”的是环绕衣壳每个五边形顶点的凹槽。与抗体发生结合的衣壳蛋白高度变异区位于这一凹槽两侧的峰顶，由于过于狭窄使得抗体不能与位于凹槽底部的残基结合，而这些残基则是衣壳与受体分子相互作用所需要的。

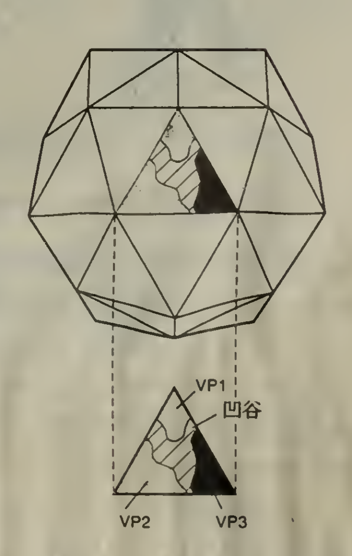


图 11-28　鼻病毒粒子表面的深裂“凹谷”

流感病毒的血细胞凝集素蛋白和神经氨酸酶是粒子表面两种可见的糖蛋白“刺突”。每个血细胞凝集素刺突由 3 个分子的聚合体构成，神经氨酸酶刺突则是由四聚体构成(图 11-30)。血细胞凝集素刺突与流感病毒和受体结合有关，它的受体为唾液酸(N-乙酰神经氨酸)，是许多不同的糖基化分子中经常存在的一种糖基。因此，这类受体作用的细胞类型特异性很低，而使流感病毒除了可以与那些可进行有效侵染细胞结合外，还可以与差异很大的不同类型细胞结合(如对红血球的凝集作用)。

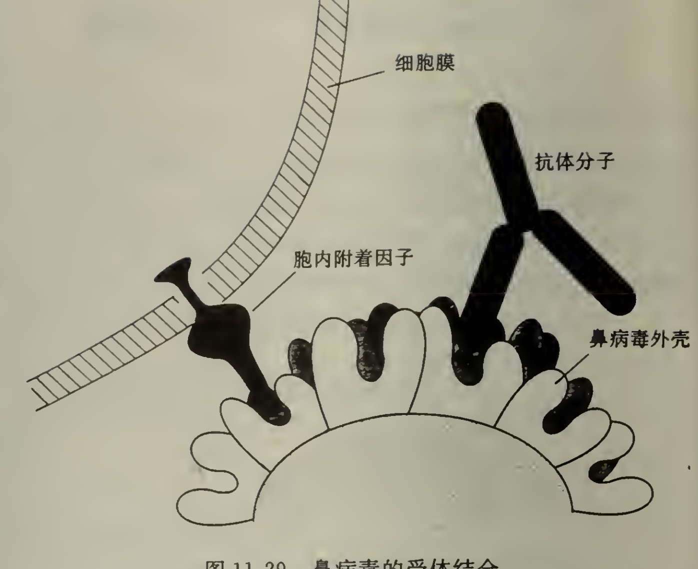


图 11-29　鼻病毒的受体结合

血细胞凝集素(HA)三聚体　神经氨酸酶(NA)四聚体
受体结合位点
13.5nm
作用位点
6nm
病毒包膜
N
基轴

图 11-30　流感病毒的糖蛋白“刺突”

在病毒的这个复制阶段中，流感病毒及副粘病毒的神经氨酸酶分子具有另一种特性。假如病毒不能很快地侵入细胞，则可能从细胞表面上脱离下来，因此对于细胞受体的吸附通常是一种可逆的过程。有些病毒具有一些特殊的脱离机制，神经氨酸酶便是其中之一。神经氨酸酶是一种脂酶，它可以将唾液酸由糖基侧链上切割下来。这对于流感病毒是非常重要的。由于受体分子的种类非常繁多，病毒可能与各种类型的细胞，甚至是细胞碎片发生一些错误的结合。与受体结合后，病毒由细胞表面的脱离经常使其发生一些变化(如病毒吸附蛋白的丢失或结构改变)，这种变化使病毒再次吸附到其它细胞上的能力降低或完全丧失。在流感病毒中，被神经氨酸酶切割下来的唾液酸残基与血细胞凝集素的活性位点结合，从而使这些分子不能再与其它受体结合。

许多情况下，受体在细胞表面的存在与否，在很大程度上决定了一种病毒的趋向性(tropism)，即病毒能够在其中进行复制的细胞种类。个别例子中，复制后期阶段细胞内的一些障碍也可影响病毒进行有效侵染的细胞种类。由此可见，复制过程的早期阶段以及病毒与寄主细胞的首次相互作用，对于病毒的致病性和侵染过程具有重要的影响。

人 γ 疱疹病毒(EBV)的复制便是由受体决定的趋性具有重要作用的一个例子。EBV 仅能感染人类细胞，在体内所侵染的细胞又分为两类：可发生有效侵染的上皮细胞和不能有效侵染的B 型淋巴细胞。对于 B 型淋巴细胞的侵染是通过细胞表面的 CR2 分子进行的。虽然上皮细胞在体内均可受到病毒以某种未知方式的侵染，但因大多数上皮细胞系不含有 CR2，在体

外也就不受感染。使用可表达 CR2 的重组 DNA 结构转染上皮细胞可以克服这种障碍，也就是使得这类细胞可以受到 EBV 的感染。

对于 HIV 的研究也得到了相似的结果。HIV 的初级受体为另一种淋巴细胞表面分子 CD4。使用重组的 CD4 表达结构转染正常情况下不能表达 CD4 的人类细胞，可以使这些细胞受到 HIV 的侵染。而用人类CD4 表达载体转染的啮齿类细胞，HIV 却不能进行有效侵染。当将 HIV 的原病毒性DNA 注入这些啮齿类细胞时，病毒便可以繁殖。这说明在细胞内部不存在对于侵染的障碍。有人推测除了 CD4 以外，还存在一种或多种对于 HIV 受体功能所需要的(人类)蛋白。另外，还有证据表明，至少在某些细胞类型中，HIV 的侵染不会因可溶性 CD4 的竞争而被阻断。因而表明，在这些细胞中，有不同的受体分子在起作用，或者至少是病毒侵入的途径不同。

在有些情况下，病毒粒子与细胞间的非特异性相互作用，能够阻断特异性受体的结合。病毒粒子虽然可能偶然地通过胞饮或吞噬作用进入细胞，但在缺少能够使病毒粒子与细胞表面保持紧密联系的某种物理作用的情况下，这种现象的发生频率是非常低的。抗体包被的病毒粒子与单核细胞或其它血细胞表面的 Fc 受体分子结合，使病毒进入细胞的过程，便是这样一个不涉及特异性病毒受体的相互作用。这种现象在很多情况下都可能发生，其中抗体对于病毒侵入的促进作用可以导致一些意想不到的结果。例如，病毒抗体的存在有时可使细胞对于病毒的摄入增加，并提高其致病性，而不是通常想象的那样对病毒发生中和作用。有人认为，这种机制在巨噬细胞和单核细胞对于 HIV 的摄入中可能也是重要的，而且这也可能是艾滋病致病过程中的一个因子。

二、侵　　入

病毒对于细胞的侵入，通常是在与细胞表面的相应受体接合后的很短时间内发生的。与吸附不同的是，对于细胞的侵入多是一种耗能过程，即：在此过程中细胞必须具有代谢活性。包括以下三种主要机制：

(1)完整病毒粒子以转位作用(translocation)穿过细胞质膜(图 11-31)。这种方式在病毒中较为少见，研究较少。肯定是由病毒衣壳中的蛋白及特异性膜受体所介导的。

(2)病毒在胞饮作用(endocytosis)下进入细胞内的液泡图(11-32)。这可能是病毒进入细胞的最常见方式。除了那些在受体结合中所需要的病毒蛋白之外，这一过程不需要其它特异性病毒蛋白，仅依赖于细胞膜上形成被有包膜的向内凹陷。由受体控制下的胞饮作用对于摄入和浓缩细胞外的大分子物质更是一种有效的过程。包膜的凹陷使得与质膜相连的液泡进入细胞质内。这些刚形成的包膜液泡保持的时间非常短，在几秒钟内便与胞内体(endosome)融合，并将所含物质释放到那些更大的液泡中。

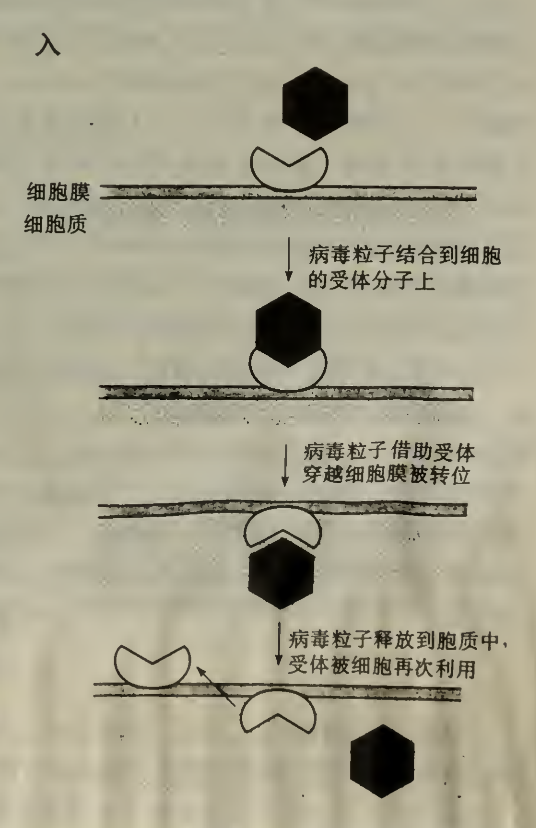


图 11-31　完整病毒粒子借助细胞的表面受体穿越胞膜进行转位

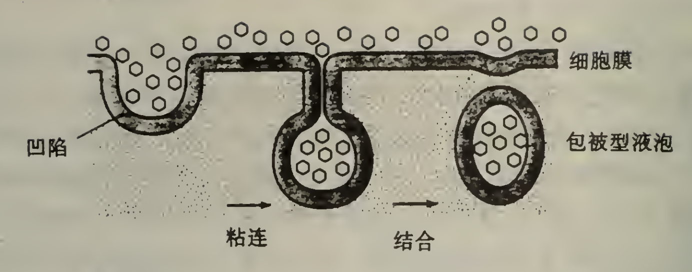




图 11-32　病毒粒子的胞饮作用和胞吐作用

此时，这些结构中所包含的病毒仍然被脂类膜与细胞质相隔离，并未真正进入细胞。随着胞内体与游离体(episome)或者是溶酶体(lysosome)的融合，由于酸化作用造成的 pH 值降低和消解酶的浓度升高，使得这些液泡内的环境逐渐恶化。因此病毒必须脱离液泡，进入细胞质，以免受到降解。这一过程的机制有多种，其中包括质膜融合及穿胞作用。病毒粒子由胞内体的释放和进入细胞质的过程与脱壳过程密切相关。

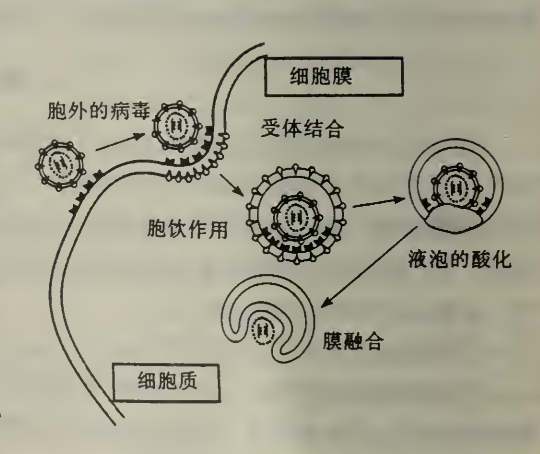


图 11-33　病毒诱导的膜融合
病毒粒子表面的特定融合蛋白受到激活后，可诱导液泡膜和病毒外膜发生融合。

(3)直接在细胞表面或胞饮进入细胞质液泡后，包膜病毒的外膜与细胞膜融合(fusion)进入细胞(图 11-33)。融合作用需要病毒包膜中的特异性融合蛋白参与，如流感病毒的细胞凝集素或逆转录病毒的 TM 糖蛋白。这些蛋白对细胞与病毒包膜的结合具有促进作用，从而使核衣壳在细胞质中直接脱落。病毒与膜的融合分为两种类型：即依赖于 pH 的融合与不依赖于 pH 的融合。

三、脱　　壳

脱壳一般是指在侵入以后，病毒衣壳被全部或部分去除，使以核蛋白复合体形式存在的病毒基因组得以外露。脱壳过程是病毒复制过程中研究较少的一个环节，认识了解也较为薄弱。

就某种意义来说，在质膜融合中病毒包膜的去除也就是脱壳过程的一部分。虽然病毒包膜与胞内体质膜间的融合是由病毒融合蛋白控制的，但这些蛋白的激活则多是由于液泡内的

低 pH 值引起蛋白构型变化，从而使原来隐蔽的融合区暴露出来。脱壳的起始可能是在胞内体内进行的，胞内体酸化引起的 pH 变化激发了这一过程，但也可能是直接在细胞质中进行的。离子载体莫能菌素或尼日利亚菌素与氯奎或氯化铵等阳离子一起使用，可以用于阻断这些液泡的酸化作用，并对脱壳究竟是发生于胞内体酸化之后（依赖 pH 值的膜融合），还是直接在细胞表面或细胞质中发生的（不依赖 pH 值的膜融合）进行测定。在小 RNA 病毒中，病毒经过胞内体对于细胞质的侵入与脱壳过程密切相关。胞内体的酸性环境导致衣壳构型发生变化，使正常病毒粒子表面并不表现的疏水区得以暴露。有人认为，这些疏水区域与胞内体膜的相互作用可以产生孔洞，使病毒基因组得以进入细胞质。

脱壳之后的产物因病毒核衣壳的结构不同而异。有些病毒可能比较简单，如小 RNA 病毒，仅有一个约 23 个氨基酸的碱性蛋白（VPg）共价结合于病毒 RNA 基因组的 5′末端。但在另一些情况下则非常复杂，如逆转录病毒的核心区即为高度有序排列的核蛋白复合体。其中除含有二倍体 RNA 外，还含有将病毒 RNA 基因组转录成为 DNA 原病毒的逆转录酶。核衣壳的结构及化学特性决定了此后的复制步骤。逆转录只能在逆转录病毒有序的核心粒子内进行，而不能在含有同样成分的反应液中进行。疱疹病毒、腺病毒及乳多空病毒的衣壳在侵入后虽然也产生了结构变化，但仍然基本保持了其完整性。这些衣壳中含有一些与细胞骨架的吸附有关的序列，经过相互作用使完整的衣壳能够进入细胞核。脱壳及核衣壳进入细胞核都是在核孔中进行的。对于呼肠孤病毒和痘病毒而言，并不发生彻底地脱壳，基因组的许多复制过程都是由病毒编码的酶所催化的，这些酶作为成熟病毒粒子的构成成分存在于细胞质内的粒子中。

四、基因组复制及基因表达

每一种病毒的复制策略都取决于其遗传物质的特性。就这一点来说，病毒可以被分为 7 种类型。这一划分方式在 1971 年由 David Baltimore 首次提出。早期仅包括 6 种类型，但现在又包括了 DNA 肝炎病毒和花椰菜花叶病毒组的基因组复制方式。对于许多病毒，尤其是含有 RNA 基因组的病毒来说，基因组的复制与遗传信息的表达是紧密联系在一起的。在以下划分的各种类型中，同时包括了复制和表达两部分内容，图 11-34 表示了它们的复制及表达策略。

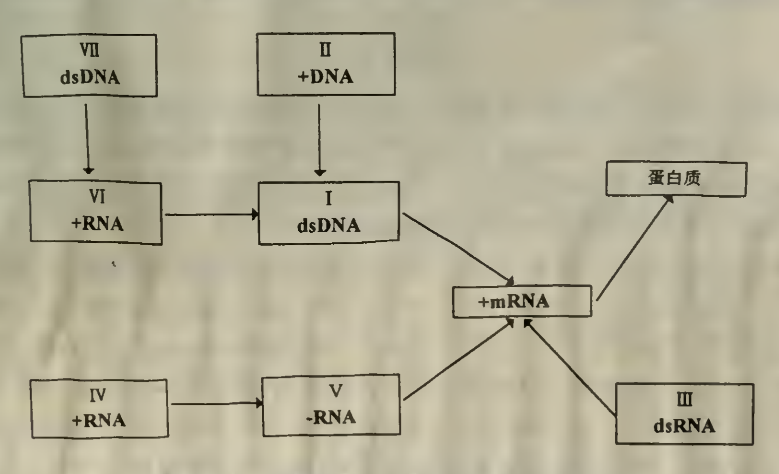


图 11-34 病毒的基因组类型及复制表达途径

Ⅰ：双链 DNA（乳多空病毒科、腺病毒科、疱疹病毒科、痘病毒科）。

这类病毒可再被分为两个类型：(1)复制过程仅限于细胞核内（乳多空病毒科、腺病毒科、疱疹病毒科），对于细胞因子具有较高的依赖性；(2)在细胞质内进行复制（痘病毒科），具备了其基因组转录及复制所有必需因子，因此对于细胞代谢具有较高的独立性。

Ⅱ：单链正意 DNA（细小病毒科）：在细胞核内进行复制，通过合成一条负意链作为正意链合成的模板。

Ⅲ：双链 RNA（呼肠孤病毒科）：病毒含有分段基因组，每个基因片段分别转录，产生各种单顺反子 mRNA。

Ⅳ：单链正意 RNA（小 RNA 病毒科、杯状病毒科、披膜病毒科、黄热病毒科、冠状病毒科、大多数植物病毒）。

可分为以下两种类型：(1)病毒含有多顺反子 mRNA（小 RNA 病毒科、杯状病毒科、黄热病毒科、马铃薯 Y 病毒组、豇豆花叶病毒组、大麦黄色花叶病毒组）所有这类病毒的 RNA 在侵染后可作为 mRNA 进行翻译，产生的多肽蛋白产物再通过切割，形成成熟蛋白；(2)病毒转录较为复杂（披膜病毒科、冠状病毒科、多数植物病毒）基因组 RNA 的产生必须至少通过两次转录方能完成。

Ⅴ：单链负意 RNA（正粘病毒科、副粘病毒科、弹状病毒科、纤丝病毒科、布尼病毒科）。

这类病毒的基因组可分为两种类型：(1)分段基因组（正粘病毒科），复制过程的第一步是由毒粒中的依赖 RNA 的 RNA 聚合酶对负链 RNA 进行转录，产生单顺反子 mRNA。这种 mRNA 也可作为下一步基因组复制的模板；(2)不分段基因组复制方式与上相同，病毒转录酶由全长病毒基因组产生每个基因的单顺反子 mRNA。

Ⅵ：具有 DNA 中间体的单链正意 RNA（逆转录病毒科）：病毒基因组虽然是由正意 RNA 构成，但在两个方面与其它病毒不同。第一，RNA 基因组为二倍体；第二，虽然 RNA 基因组并不能作为 mRNA，但可作为模板逆转录产生 DNA。

Ⅶ：具有 RNA 中间体的双链 DNA（DNA 肝炎病毒科、花椰菜花叶病毒组）：这类病毒虽然也存在逆转录，但与逆转录病毒不同的是，它的发生是在成熟过程中的病毒粒子内进行的。侵染一个新的细胞后，首先要对带有缺口的基因组进行修补然后才能转录。

五、装　　配

装配过程就是构成病毒粒子所必需的所有组分，在细胞内特定场所的集合，形成病毒粒子的基本结构。装配场所取决于病毒在细胞内的复制地点和不同的释放机制。例如，小 RNA 病毒、痘病毒及呼肠孤病毒在细胞质中复制，腺病毒、乳多空病毒及细小病毒在细胞核中复制，而逆转录病毒则是在细胞膜的内侧表面进行复制的。

在复制的早期阶段，有时无法将装配、成熟和释放过程作为不同的独立阶段进行研究。但在任何情况下，装配场所却对这些过程具有重要的影响。在绝大多数情况下，细胞膜都被用来附着病毒蛋白并起始装配过程。目前一般认为，随着细胞内病毒蛋白和基因组分子浓度的提高，在达到一定水平时便可启动装配过程。许多由病毒合成的结构成分，在细胞内的局部区域可以达到很高的水平。这种浓缩形式叫做内含体(inclusion body)，可在光学显微镜下观察到，也是许多病毒导致细胞侵染后期阶段的一种特征。内含体的大小和在侵染细胞内的位置，对于某一病毒来说也常具有特征性。如：1903 年由 Adelchi Negri 首次发现的，狂犬病毒

感染所产生的大型核周“涅格里氏体”。另外，病毒结构成分的局部浓度也可因膜相关蛋白的相互作用而提高。这种机制对于借助出芽由细胞内释放的包膜病毒来说尤为重要。

仅仅由衣壳亚基间的相互作用所决定，并遵循对称性规律形成病毒粒子，是一种较为简单的过程。在另外一些情况下，病毒的装配则是一个非常复杂、包括了许多步骤的过程。它不仅涉及病毒结构蛋白，同时涉及一些来自于病毒或细胞，作为模板指导毒粒装配的骨架蛋白。病毒基因组的衣壳包装可以是在粒子装配的早期阶段（如：许多螺旋对称病毒衣壳蛋白与基因组的结合），也可以是在装配的后期阶段。在后一种情况下，基因组是被装配放入一个几乎完整的蛋白外壳中的。

六、成　　熟

经过这个阶段，病毒方能具备侵染性。成熟过程通常涉及病毒粒子的结构变化，即通过对衣壳蛋白的特异性切割而形成成熟蛋白产物或在装配中使蛋白发生构型变化。这种变化经常导致衣壳结构的本质性改变，通过对未成熟病毒及成熟病毒粒子的抗原性差异比较可以确定这些改变，在有些情况下抗原性可能完全不同（如：小 RNA 病毒）。另外，核蛋白与病毒基因组结合所引起的内部结构的改变也经常导致上述变化。

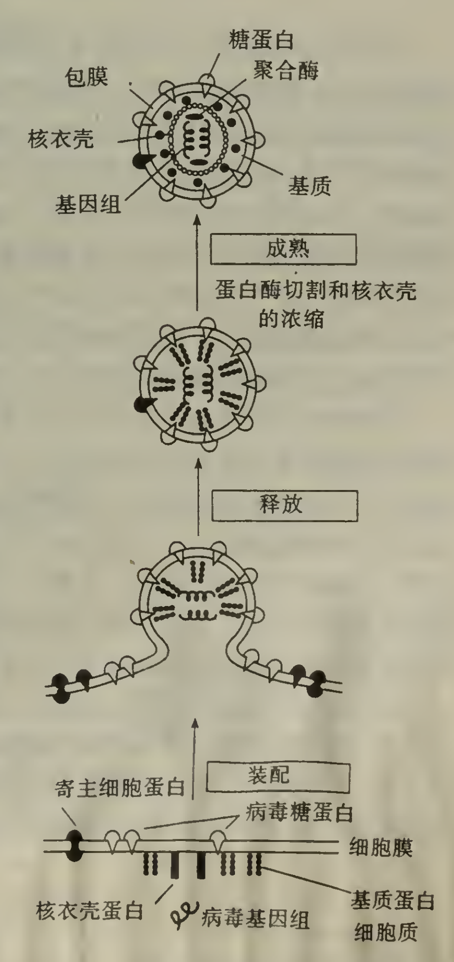


图 11-35　病毒通过“出芽”方式释放

虽然在有些情况下细胞中的酶或病毒与胞内酶的混合物也会参与成熟过程，但由于细胞内蛋白酶的底物特异性较低，容易将衣壳蛋白彻底降解，它的作用存在很多弊端。因此，病毒编码的蛋白酶在成熟中发挥着主要作用，它们通常具有针对特定氨基酸序列和结构的高度特异性，多数情况下仅对一个大的复杂病毒衣壳中的某个特定肽链进行切割。另一方面，这些病毒来源蛋白酶的功能，由于在装配中被包装进入病毒颗粒而受到调控。仅在由于衣壳构型、局部疏水环境、衣壳内部 pH 或金属离子浓度变化时，它们才能与目标序列紧密接触，并具有活性。在成熟过程中受到严格调控的酶类中，逆转录病毒的蛋白酶便是一个很好的例证。逆转录病毒的核心粒子由 *gag* 基因编码的蛋白所构成，其蛋白酶在出芽释放之前被包被于核心粒子之中。在出芽过程的某个阶段（因不同逆转录病毒而异），蛋白酶对 gag 蛋白前体进行切割，产生各种成熟蛋白，即成熟病毒粒子的衣壳蛋白、核衣壳及基质蛋白（图 11-35）。

在成熟过程中，并非所有的蛋白酶切割都是受到这种严格调控的。天然的流感病毒血细胞凝集素翻译后需进行修饰（在高尔基体上糖基化），并在此阶段表现出受体结合活性。然而，这种蛋白必须被切割成为两个片段（HA_1 和 HA_2），才能在侵染中使膜发生融合。在释放到细胞表面之前，切割作用发生在病毒出芽进入的分泌液泡中，细胞

内的类胰蛋白酶与此过程有关。金刚胺是一种对A型流感病毒具有抑制活性的药剂。虽然这种试剂的作用机理是较为复杂的，尚未全部了解，但它肯定可以阻断细胞膜的离子通道。这种药剂的作用靶点是流感病毒的基质蛋白(M_2)，但对于药剂的抗性却是在血细胞凝集素基因上。某些流感病毒株系的复制可以在细胞侵入阶段受到抑制，而另一些则发生在成熟期。药剂的这种不同作用是由于在经处理的细胞内，病毒不能对液泡内的低pH做出反应(这种功能多由M_2基因产物所控制)，因此也就不能在成熟过程中对血细胞凝集素进行切割。与此类似，逆转录病毒包膜糖蛋白也需要切割方能具有表面蛋白(SU)和穿膜蛋白(TM)的活性。关于细胞内的酶类参与这一过程的机制所知甚少。

如上所述，对某些病毒而言，装配和成熟过程都是在细胞内进行的，也是难以区分的，但另一些病毒的成熟则仅发生于病毒粒子由细胞中释放以后。在任何一种情况下，成熟过程都使得病毒粒子具备了对于细胞的侵染性。

七、释　　放

对于溶源性病毒(如大多数非包膜病毒)来说，释放过程是比较简单的，即细胞破碎而释放出病毒。有包膜病毒释放之前则需要在通过细胞膜向外出芽或进入胞内液泡时形成脂类膜。病毒的包膜蛋白是在粒子释放过程中装配到外膜上的，这个过程称为出芽。病毒粒子以这种方式释放可能对细胞造成严重的伤害(如弹状病毒和疱疹病毒)，也可能不造成伤害(如逆转录病毒)。但无论在哪种情况下，这一过程都是由病毒所决定的，即衣壳蛋白对于细胞膜内壁的物理作用使病毒粒子得以穿过膜结构(图11-35)。对于以出芽方式释放的病毒来说，装配、成熟及释放，是一些同时发生的过程。与病毒出芽有关的膜的类型是由相应病毒决定的，即可以是细胞的外膜(如逆转录病毒和披膜病毒)，也可以是细胞质膜(如正粘病毒、副粘病毒、布尼病毒、冠状病毒、弹状病毒及肝炎DNA病毒)或核膜(如疱疹病毒)。

某些病毒的膜蛋白不仅作用于释放过程，也可以在复制的初始阶段具有功能。流感病毒的神经氨酸酶蛋白便是这种表现的一个很好例子。神经氨酸酶不仅能够通过血细胞凝集素使病毒粒子与细胞的结合逆转，它还被认为在防止流感病毒粒子聚集及病毒释放中具有重要作用。除了特异性蛋白以外，具有出芽功能的病毒还通过在装配—成熟—释放过程中精确的时间顺序来解决释放中的问题。虽然不太可能通过生物化学分析将这些阶段区别开来，但是这些过程在空间位置上却存在着一些微小的差别。另外，在复制过程后期的阶段，伴随着病毒粒子和包膜内所发生的许多构象变化，有些病毒仍然可以进行复制。

总之，所有类型的病毒复制包括了3个大的阶段，即侵染的起始、基因组的复制和表达，以及最后成熟病毒粒子由侵染细胞中的释放。虽然不同病毒在这一过程中都存在着一些表现形式不同的相似阶段，但它们在复制过程的细节上又存在着许多差异。这是由于不同类型寄主细胞的生物学及病毒基因组的特性所决定的。

第五节　病毒基因的表达

一、遗传信息的表达

病毒与寄主细胞间最主要的相互作用，就是病毒要在细胞器官中进行核酸及蛋白质的合

成。因此，病毒的复制过程也受到寄主基因表达的严格控制。就真核细胞和原核细胞来说，各自有着完全不同的表达调控机制，这些差异不可避免地对利用它们作为寄主的病毒产生影响。另外，与最简单的细胞相比，更为精炼的病毒基因组在这一过程中也表现出很强的调控能力。细胞可以利用丰富的遗传信息，通过各种不同的复杂机制来对基因的表达进行调控。而病毒则要在有限的遗传资源中，通过严格的定量和时间、空间上高度特异性，来实现对表达的控制。

病毒需要通过一系列方式来弥补遗传资源的不足。其中包括，能够高效启动或阻遏基因表达的正、负调控信号；高度精炼，并经常含有重叠阅读框架的基因组；常常包含在其它基因中的调控信号；以及由一条 mRNA 合成多种蛋白的不同策略。基因表达包括了由各种信号参与的调节过程，这些信号中有些以顺式方式作用于相邻的基因(*cis*)，有的则可能通过其产物对相邻或不相邻的调控位点发生反式作用(trans)。例如，转录启动子便是顺式作用序列，它与所调控的基因相邻。而细胞内与核酸特定序列结合的“转录子”蛋白则是反式作用因子。较小的基因组和所具有的精确调控机制构成了病毒遗传调控的基本模式。

二、λ 噬菌体的表达调控

对 λ 噬菌体基因组研究较其它病毒更为深入，目前已经得到了一些有关机制的结果。如阻遏蛋白在溶源(lysogeny)复制与裂解(lytic)复制之间的调控作用，以及噬菌体编码的顺式调控因子在抗转录终止中的作用。主要以前噬菌体(prophage)形式存在于大肠杆菌中的 λ 噬菌体即可以进行溶源性生长，也可以进行裂解生长。图 11-36 所示为 λ 噬菌体的简单基因图谱。除编码病毒外壳蛋白的头、尾部分的结构基因外，在 λ 噬菌体生活史中，以下元件在基因调控中的功能更为重要。

在新感染细胞中，*N* 基因和 *cro* 基因分别由 P_L 和 P_R 转录(图 11-37)。N 蛋白使 RNA 聚合酶对多个噬菌体基因进行转录，包括那些与 DNA 重组、前噬菌体整合有关的基因以及 *cII*

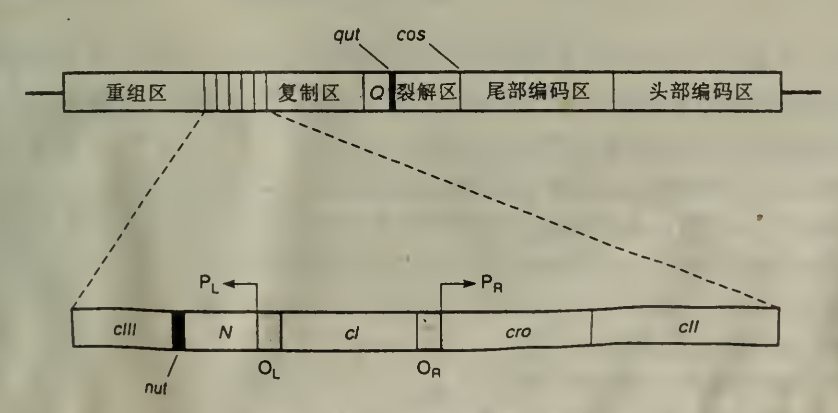


图 11-36 简化的 λ 噬菌体遗传图谱

P_L：包括 *N* 和 *cIII* 基因在内的 λ 基因组左侧区域转录启动子。O_L：基因组中的一小段非编码区(约 50 bp)，位于 *cI* 和 *N* 基因之间，与 P_L 相邻。P_R：包括 *cro*、*cII* 和编码结构蛋白的基因在内的 λ 基因组右侧区域转录启动子。O_R：基因组中的一小段非编码区(约 50 bp)，位于 *cI* 和 *cro* 基因之间，与 P_R 相邻。*cI*：以其自身的启动子转录，编码一种 236 个氨基酸的阻遏蛋白。此蛋白与 O_R 基因结合后将阻止 *cro* 的转录，但仍允许 *cI* 转录；与 O_L 基因结合后则阻碍 *N* 基因及位于基因组左侧的其它基因的转录。*cII* 和 *cIII*：编码与基因组结合的激活蛋白，具有增强 *cI* 基因转录的作用。*cro*：编码一种 66 个氨基酸的蛋白，此蛋白与 O_R 基因结合后可以防止阻遏物与它的结合。*N*：编码一种反终止子蛋白，通过对寄主细胞 RNA 聚合酶的活性进行修饰，保证了 P_L 和 P_R 的高效率转录，作用与 ρ-因子类似。*Q*：是与 *N* 基因相似的反终止子，但它只允许从 P_R 起始的延伸转录。

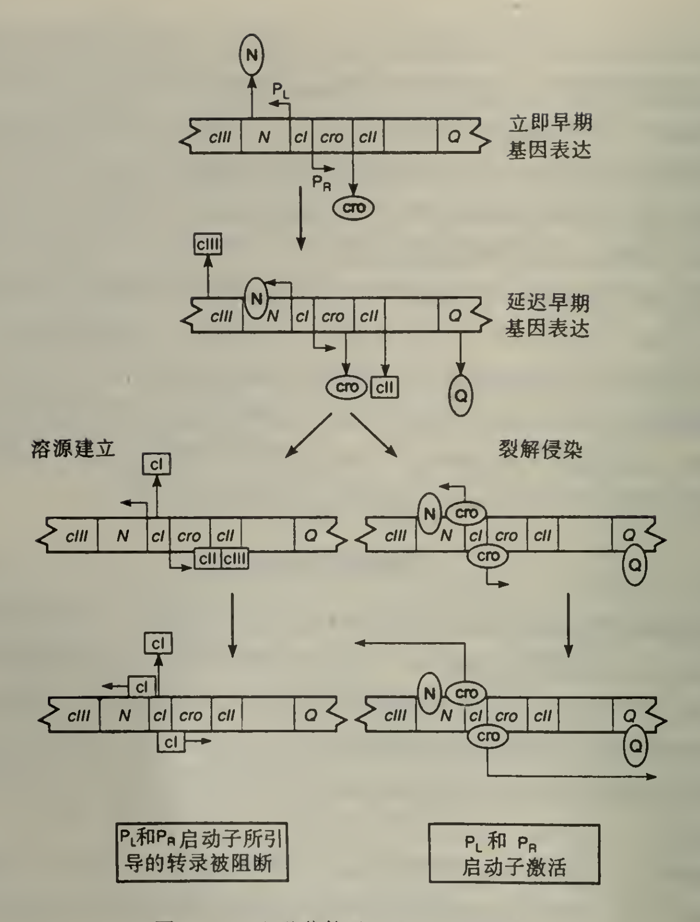


图 11-37 λ 噬菌体基因组的表达调控

和 *cIII* 基因。N 蛋白作为转录正调控子起作用。当没有 N 蛋白时，RNA 聚合酶全酶分别在位于 *N* 和 *Q* 基因末端的 *nut* 和 *qut* 位点处中止。RNA 聚合酶-N 蛋白复合物是由 P_L 和 P_R 引导下的转录能够完成，而 RNA 聚合酶-Q 蛋白复合物则仅保证了 P_R 引导下的转录。随着细胞中 cII 和 cIII 蛋白水平的增加，*cI* 阻遏子基因由自己的启动子引导开始转录。

通常情况下，cII 蛋白会不断地受到细胞中蛋白酶的降解，当 cII 蛋白水平保持在某个临界水平以下时，由 P_L 和 P_R 引导的转录将持续进行，噬菌体进行增殖性复制，最终使细胞裂解并释放出噬菌体颗粒。但在极少数情况下，cII 蛋白的浓度会积累升高，从而促进了 *cI* 基因的转录，并导致细胞内 cI 阻遏蛋白水平的上升。这种阻遏蛋白与 O_R 和 O_L 区域结合，阻碍了除自身以外的所有噬菌体基因的转录。但是过高浓度的 cI 阻遏蛋白也会与 O_R 区域的左端结合，从而阻止 *cI* 的转录（图 11-38）。所以，cI 蛋白的水平受到一种负反馈机制的自动调节，使被感染细胞保持稳定的溶源状态。

这些溶源细胞又是如何转换进入增殖性裂解复制循环的呢？生理压力，以及尤其是紫外线对细胞的辐射，可以诱导寄主细胞产生一种称作 RecA 的蛋白质。这种蛋白可以降解 cI 阻遏蛋白，它的正常功能是诱导细胞的某些基因表达，使细胞能够在变化的环境条件下适应、存活。虽然 RecA 蛋白本身并不足以阻止细胞重新进入溶源状态，但当 cI 阻遏蛋白不能与 O_R 结合时，*cro* 基因便可以被 P_R 引导开始转录。所产生的 cro 蛋白也可以与 O_R 区域结合。但与

主要结合于 O_R 区域右端的 cI 蛋白不同的是，cro 蛋白优先与 O_R 区域的左端结合，阻断了 *cI* 基因的转录，并通过一种正反馈机制促进其自身的转录，从而使噬菌体保持裂解复制，不能恢复到溶源状态。

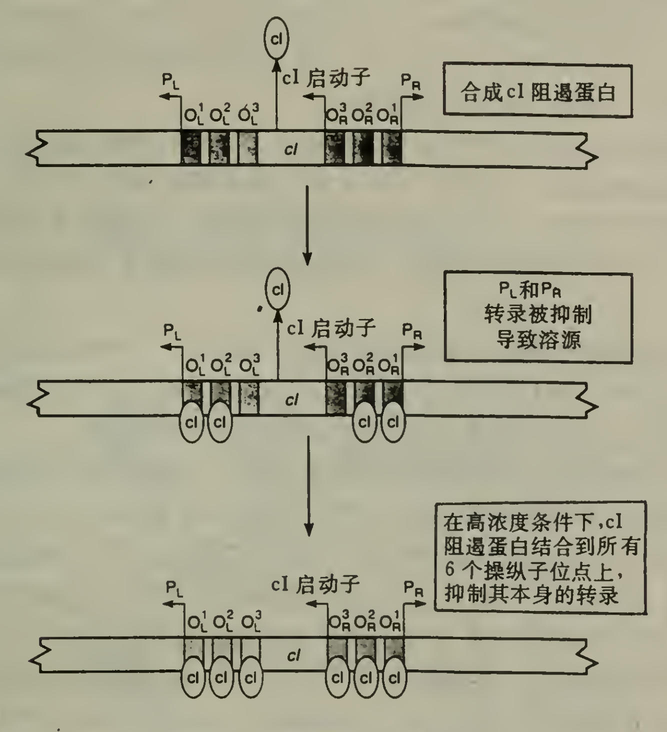


图 11-38 λ 噬菌体的溶源调控

三、病毒基因组的编码策略

在上一节中，根据结构将病毒基因组分为 7 种类型。由于病毒基因组的复制和表达是密不可分的(尤其是 RNA 病毒)，下面再对上述 7 类病毒基因组的表达方式逐一进行讨论。

（一）双链 DNA 病毒(腺病毒科、疱疹病毒科、痘病毒科及乳多空病毒科)

如上节所述，这些病毒都与细胞双链 DNA 相似的基因组，基本上是采用与细胞基因相同的机制进行转录。就这一点来说，可以说它们是相似的，但是每一类病毒对寄主细胞器官的依赖程度却存在很大差异。

1. 乳多空病毒　乳多空病毒的复制和基因表达均高度依赖于细胞器官。其中多瘤病毒和乳头瘤病毒都编码一些反式作用因子(T-抗原)，来起始转录及基因组复制。尤其是乳头瘤病毒的复制更要依赖于寄主细胞，虽然这些病毒能够编码几种反式调控蛋白，但仅能在分化末期的角质化细胞中复制，而不能在其它类型的细胞中复制。

2. 腺病毒　腺病毒的转录同样高度依赖于细胞器官。但是它们采用不同的机制来对病毒基因的表达进行调控。在这些机制中，包括了 E1A 蛋白等反式作用转录激活子，以及通过 mRNA 的拼接和病毒编码的 *VA* RNA 对表达进行转录后调控。腺病毒对于细胞的感染分为早、晚两个阶段，后一个阶段与基因组复制同时开始，在不同腺病毒之间差异较小。

3. 疱疹病毒　这类病毒比上述各种病毒对细胞中酶类的依赖程度较低。它们编码多种与

DNA 代谢有关的酶(如胸苷激酶)和一些反式作用因子，后者对病毒基因的表达时间进行调控，从而控制侵染的进程。这些病毒的基因组较大，也较复杂，它的转录是以串联方式按顺序调控的。在寄主细胞 RNA 聚合酶 II 的参与下，基因组转录后至少可产生 50 种病毒编码的蛋白。其中涉及以下 3 种类型的 mRNAs：①α-立即早期(IE)mRNAs，编码 5 种对于病毒转录的反式作用调节子；②β-(延迟的)早期 mRNAs，编码非结构调控蛋白和某些较次要的结构蛋白；③γ-晚期 mRNAs，编码主要的结构蛋白。

疱疹病毒的基因表达是受到严格而协调控制的。即：①如果蛋白翻译在感染后很快受到阻断(如用放线菌酮处理细胞)，早期 mRNAs 立即在细胞核中积累，但不进一步转录产生病毒 mRNA；②早期基因产物的合成会终止立即早期产物的形成并起始基因组的复制；③某些晚期结构蛋白的产生不依赖于基因组复制，而其它蛋白则只能在基因组复制后，由新复制的 DNA 转录产生。

基因组复制的起始同时需要立即早期蛋白和早期蛋白。一种病毒编码的依赖于 DNA 的 DNA 聚合酶(DdDp)和一种 DNA 结合蛋白，以及多种影响细胞生化代谢的酶(如胸苷激酶)共同参与基因组复制。所有这些蛋白质的产生都是受到严格调控的。

4. 痘病毒　在痘病毒中，基因组的复制和基因表达，除需要寄主细胞核糖体外，几乎独立于细胞代谢。痘病毒基因组编码许多与 DNA 代谢、病毒基因转录和 mRNA 转录后修饰有关的酶，其中许多种类都包含在病毒颗粒中。这使得在细胞质中进行的基因组转录、复制，几乎全部都在病毒自身控制之下，而上述其它类型的病毒则需要在细胞核中复制。基因表达是由与病毒粒子核心结合的酶指导的，可分为两个差别较大的阶段。即：

(1)约占痘病毒基因组 50%的早期基因，可以在部分脱壳的核心粒子内，于基因组开始复制以前进行表达，产生具有 5′-帽子、3′-polyA 结构，但没有剪接的 mRNAs。

(2)晚期基因于基因组开始复制以后，在细胞质中进行表达。但这些基因的表达也是依赖于病毒编码的转录蛋白，而不是位于细胞核内的细胞转录蛋白。如同疱疹病毒一样，痘病毒晚期基因启动子的激活也依赖于 DNA 的早期复制。

(二)单链正意 DNA 病毒(细小病毒科和双联病毒组)

由于细小病毒的基因组非常小，不足以编码各种必需的生化产物，所以不管是自主型还是依赖于辅助病毒的细小病毒，它们的基因表达和基因组复制都在很大程度上依赖于来源于外部的帮助。它们表达和复制都是利用寄主细胞核所具有的正常功能来进行的。植物双联病毒的基因组结构也属于这种类型，虽然它们的基因组表达与细小病毒相差较大，但是也是高度依赖于寄主细胞的功能。

细小病毒科的复制缺陷型依赖病毒属(*Dependovirus*)成员，完全依赖于腺病毒或疱疹病毒的双重感染来提供复制必需的辅助功能。作为辅助子的 E1A 等腺病毒基因是早期转录的调控基因，但对晚期结构基因没有作用。最近的研究结果表明，用紫外线、放线菌酮或某些致癌物处理细胞可以取代辅助病毒的作用。因此，所需要的辅助作用可能仅是对细胞环境的改变，而不是一种可能影响缺陷型细小病毒基因组转录的特定病毒蛋白。

(三)双链 RNA 病毒(呼肠孤病毒科，多数真菌病毒)

呼肠孤病毒在寄主细胞质中复制，并具有多分体基因组。在侵染过程中，在病毒颗粒中含有的一种病毒特异性 RNA 聚合酶作用下，以全保留方式由每个基因组片段分别产生独立的单顺反子 mRNA(图 11-39)。即仅以核心粒子内的双链基因组中的负意链为模板，合成产生

mRNA，释放后由这些 mRNA 翻译产生复制所需要的病毒非结构蛋白.原有双链基因组仍然保持完整。在侵染过程的后期，单链正意 RNA 也被病毒编码的复制酶复制产生负意 RNA 链，并与正意链共同形成新的双链基因组片段。在基因组复制结束之后，由新形成的基因组片段转录产生编码结构蛋白的晚期 mRNA，并从病毒核心粒子中释放翻译，进而形成成熟的病毒颗粒。

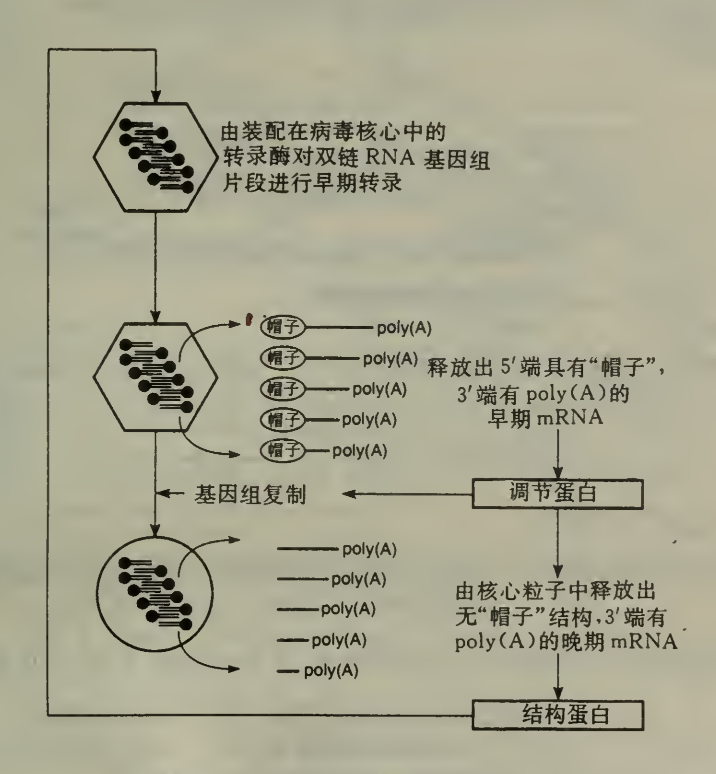


图 11-39　呼肠孤病毒基因组的表达

（四）单链正意 RNA 病毒

这一类型中包括了多种动物（杯状病毒科、冠状病毒科、黄热病毒科、小 RNA 病毒科和披膜病毒科）和植物病毒（雀麦花叶病毒组、香石竹斑驳病毒科、豇豆花叶病毒科、黄瓜花叶病毒科、石竹病毒科、真菌传杆状病毒科、大麦病毒科、烟草线条病毒科、大麦黄矮病毒科；线虫传多角体病毒科、马铃薯 X 病毒科、马铃薯 Y 病毒科、南方菜豆花叶病毒科、烟草花叶病毒科、烟草脆裂病毒科、番茄丛矮病毒科和芫菁花叶病毒科）。是病毒种类及科的数目最多的一类病毒基因组。这些病毒基因组可作为 mRNA，在感染寄主细胞后立即进行翻译。由于包括的病毒种类繁多，这一类基因组在基因表达和基因组复制的调控策略上表现出极大的差异性。这类病毒在广义上可分为以下两个类群：

（1）整个基因组首先编码一种多蛋白（polyprotein），然后被病毒编码的特异性蛋白酶切割成为前体蛋白和各种成熟的蛋白。由于不同的切割可以从一种前体产生性质各不相同的几种蛋白，所以这些切割对遗传信息的表达是一种非常精细的调控方式，例如，小 RNA 病毒和马铃薯 Y 病毒（图 11-40 A）。某些含有多分体基因组的植物病毒也采用了一种非常相似的策略，从每个基因组片段分别产生多蛋白，来对基因的表达进行调控。基因组的结构与小 RNA 病毒

非常相似的豇豆花叶病毒便是一个极好的例子(图 11-40 B)。

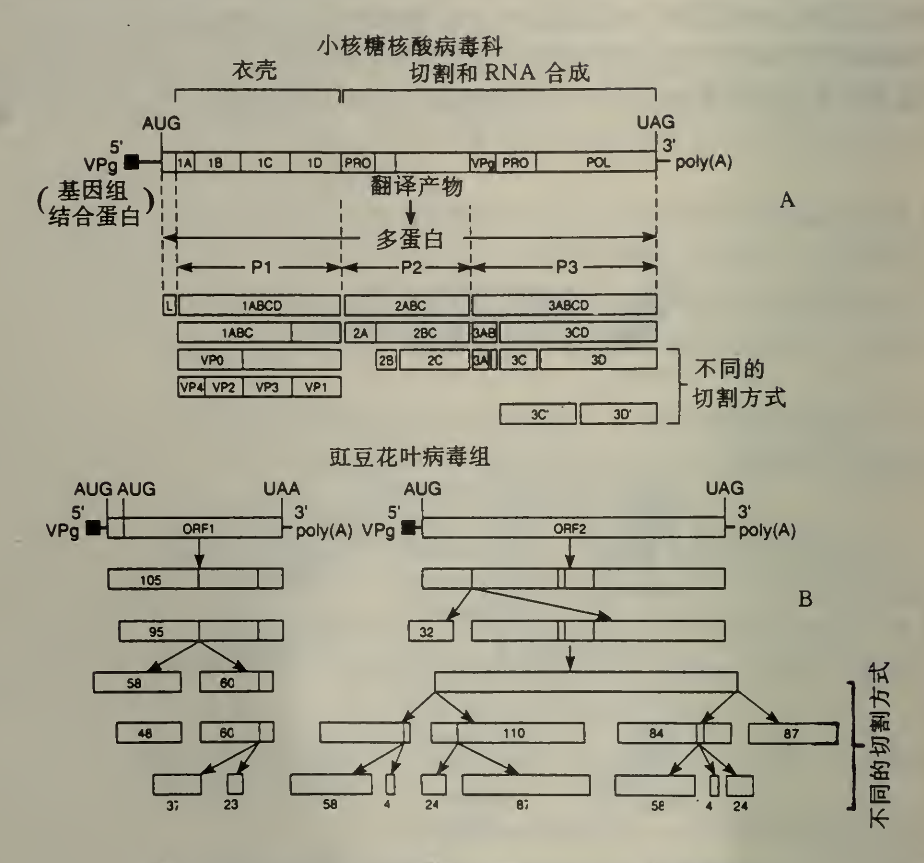


图 11-40　以多蛋白为初始翻译产物的正链 RNA 病毒基因组

B 中数字后的单位为 kD。

(2)产生亚基因组 mRNA，基因组进行两次或更多次数的翻译循环。这一策略使复制过程在时间上能够分隔开来，实现必要的“早期”和“晚期”阶段。在早期阶段中，产生包括一种病毒复制酶在内的非结构蛋白，而在以后的晚期阶段中则产生一些结构蛋白(图 11-41)。在以上每个复制阶段中所产生的蛋白质，都有可能是通过对多蛋白水解加工而来的，但与上一种类型不同的是，这些蛋白仅占全部基因组遗传信息的一部分，而不是全部。蛋白质的酶解又可以对在每个复制阶段中所产生的不同蛋白的比例进行调控(如披膜病毒和马铃薯 X 病毒)。除了对蛋白质酶解以外，某些病毒还采用了另外的策略来产生不同的蛋白，如产生亚基因组 mRNA、对翻译终止密码的通读(read-through)(烟草花叶病毒)、在特定位点上发生的核糖体移码(红三叶草坏死斑驳病毒)、或使用相互重叠的读码框架(番茄丛矮病毒)。

总之，所有的这一类病毒的基因表达调控都通过以下机制来实现的，也就是病毒编码不同蛋白的比例以及产生这些蛋白的复制阶段。与前面谈到的两类 DNA 病毒基因组相比，这些机制在很大程度上独立于它们的寄主细胞。

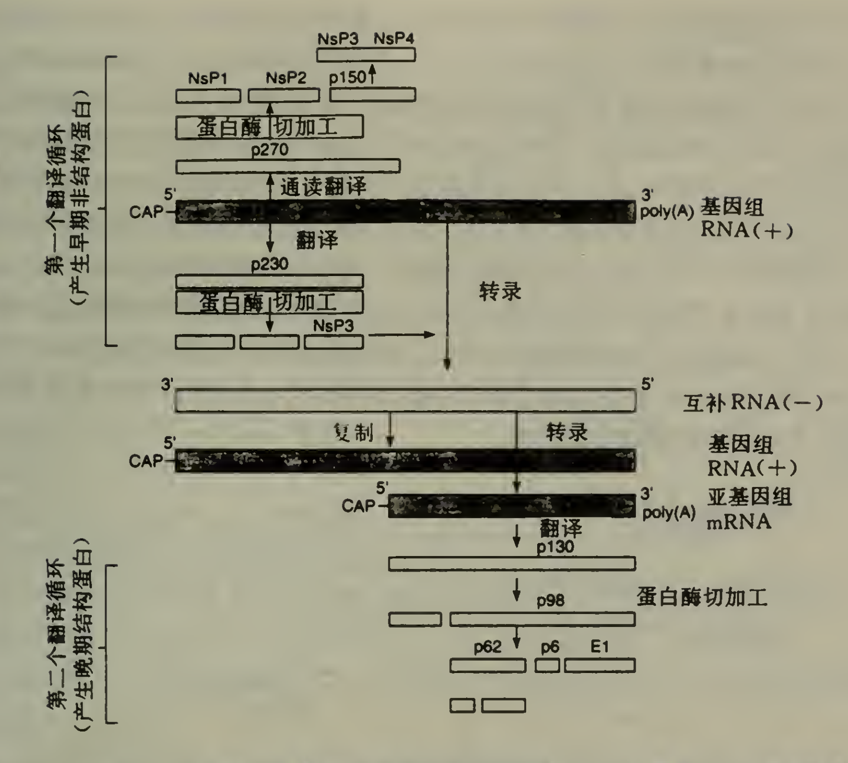


图 11-41　通过两个相互独立的翻译循环进行表达的正链 RNA 病毒基因组（披膜病毒）在侵染后期产生亚基因组 mRNA。

（五）单链负意 RNA 病毒（正粘病毒科、副粘病毒科、弹状病毒科、纤丝病毒科和布尼病毒科）

正粘病毒分段基因组的复制，第一步是负意 RNA 在与病毒粒子结合的 RdRp 作用下，转录产生单顺反子 mRNA，这种 mRNA 也是下一步基因组复制的模板（图 11-42）。在所有的负意 RNA 病毒中，由于寄主细胞内不含有任何能够破译和复制这类 RNA 基因组的酶，所以病毒特异性的转录酶或复制酶必须包装在病毒的核衣壳中。

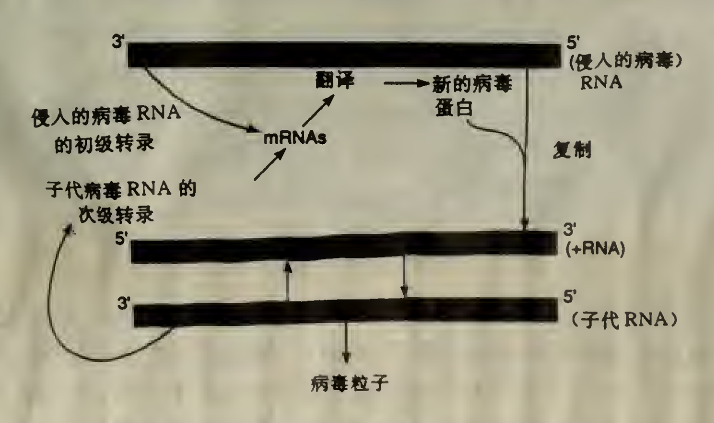


图 11-42　负链 RNA 病毒基因组表达示意图

具有不分段基因组的其它病毒科，同样也产生单顺反子 mRNA。但是，对于不分段基因组来说，mRNA 必须来源于一个长的负意 RNA 分子。如何实现这一点的确切机制尚不清楚。可能是由一个全长基因组转录子，经过转录后切割形成不同的 mRNA。更有可能的是，这些 mRNA 是通过一种转录的终止一起始机制分别产生的，而这种机制又受到存在于每个病毒基因之间的保守序列的调控。由于所有的病毒在产生每个病毒颗粒时需要的结构蛋白（如核衣壳蛋白）要比非结构蛋白（如聚合酶）多得多，所以，除了上述基因表达模式以外，在转录过程中和转录以后都还要对不同蛋白的比例进行调控。例如，在副粘病毒的转录中，存在一种对于病毒基因组由 3′端到 5′端转录的明显极向性，从而使编码结构蛋白的基因组 3′端 mRNA 合成远远多于位于 5′端的非结构蛋白编码区域，3′端基因翻译产生的结构蛋白和 5′端基因翻译产生的非结构蛋白的相对量可以相互调节（图 11-43）。与此相似，在产生的每个单顺反子 mRNA之间，也会存在翻译效率的差异。

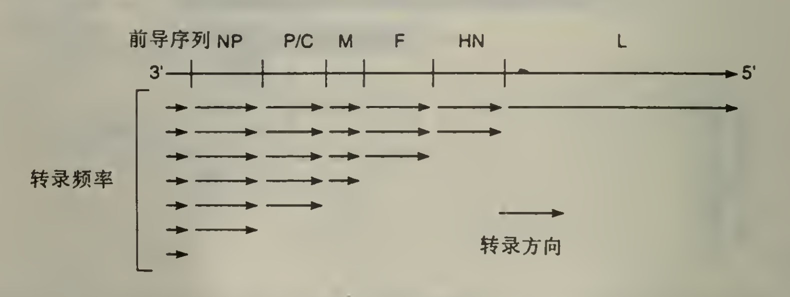


图 11-43　副粘病毒基因组的“转录极性”

（六）以 DNA 为中间体的单链正意 RNA 病毒（逆转录病毒科）

逆转录病毒的 RNA 基因组不具有 mRNA 的功能，而是作为逆转录模板的产生 DNA。DNA 原病毒（provirus）一旦整合到寄主细胞的基因组中，其所有的功能便都在寄主细胞的控制之下，并完全像其它细胞基因一样受到转录。但也有一些逆转录病毒具有若干转录及转录后调节机制，来对它们的遗传信息的表达进行调控。

（七）以 RNA 为中间体的双链 DNA 病毒（DNA 肝炎病毒科、花椰菜花叶病毒）

这些病毒基因组的表达较为复杂，所知甚少。DNA 肝炎病毒含有若干重叠的阅读框架，是一种高度精炼、压缩了最大数量编码信息的基因组。*X* 基因编码一种反式转录激活子，但它在感染中的作用尚不清楚。由各自独立的启动子至少产生两种 mRNA，每一种 mRNA 都编码几种蛋白，其中较大的一种又是在病毒颗粒形成中的逆转录模板。花椰菜花叶病毒基因组的表达也较复杂。两种主要 mRNA 的产生与 DNA 肝炎病毒相似。每一种 mRNA 都编码几种多肽，较大的一种在病毒基因组的形成中也作为逆转录的模板。

四、表达的转录调控

猴病毒 40（SV40）是乳多空病毒科成员，对于它的基因组及其相关的细胞基因组研究较为深入。SV40 基因组编码两种肿瘤抗原（T-抗原），根据蛋白的分子量分别称为大 T-抗原和小 T-抗原（图 11-44）。SV40 双链 DNA 基因组在寄主细胞核中复制。基因组在寄主细胞 RNA 聚合酶 Ⅱ 的作用下进行转录，同时大 T-抗原在病毒基因组的转录调控中发挥着重要作用。小

T-抗原对于病毒的复制不是必需的，它使病毒 DNA 在细胞核中积累。两种蛋白都含有可以使病毒在细胞质中合成后，向细胞核内积累的核导向信号(nuclear localization signals)。

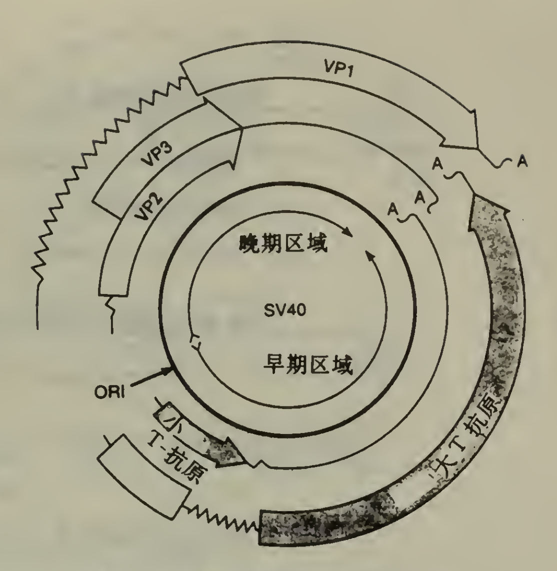


图 11-44　猴病毒 40 的基因组结构和可能的蛋白质编码区

在对细胞侵染后不久，便由早期启动子引导表达早期 mRNA。在这个启动子中包含一段由 72 bp 构成的重复序列，具有很强的转录增强子作用，激活启动子在新感染的细胞中的功能(图 11-45)。所合成的早期蛋白便是上述两种 T-抗原。随着大 T-抗原在细胞核中的浓度上升，大 T-抗原蛋白与基因组起始区的结合，使早期基因的转录受到抑制，阻断了由早期启动子引导的转录，使其进入侵染的晚期阶段。在 DNA 的复制开始以后，晚期启动子便引导起始晚期基因的转录，合成结构蛋白 VP1、VP2 和 VP3。因此，SV40 T-抗原在基因组转录调控中的作用就像一个开关。

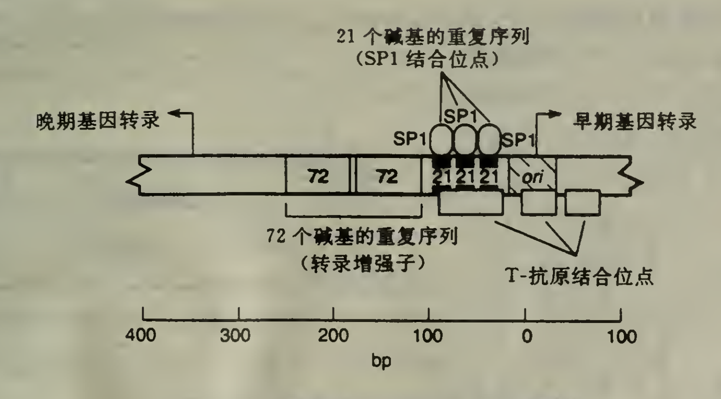


图 11-45　猴病毒 40 基因组的转录调控

在病毒转录调控研究中，另一类较为深入的工作是在人类逆转录病毒中开展的，即人的嗜 T 细胞白血病病毒(HTLV)及艾滋病毒(HIV)。逆转录病毒的 RNA 基因组经过逆转录形成整合型的 DNA 原病毒(provirus)。通过 DNase I 足迹蛋白和凝胶移位实验分析表明，在这些病毒的长末端重复序列(LTR)中，存在许多细胞转录因子的结合位点(图 11-46)。在 LTR 的 U3 区域中，由远端序列(如 NKκB 和 SP1 结合位点)和近端元件(如 TATA 盒)一起构成了功能性转录启动子。但这些启动子本身的活性较弱，只能在 RNA 聚合酶 I 的作用下对原病毒基因组进行有限地转录。

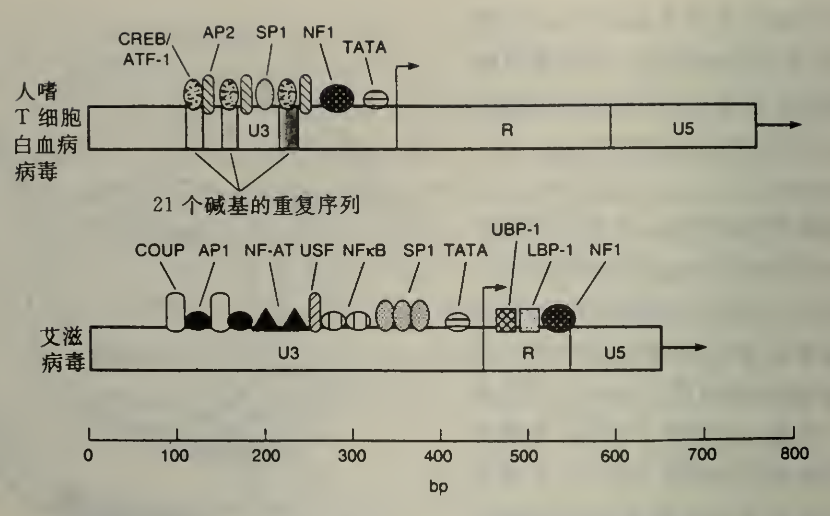


图 11-46　与逆转录病毒 LTR 相互作用的胞内转录因子

HTLV 编码的 tax 蛋白和 HIV 编码的 tat 蛋白都是转录过程中反式作用的正调控蛋白(图 11-47)。借助于一种比原来启动子效率高 50～100 倍的因子，这些蛋白可以增强由病毒 LTR 引导的转录。目前，这些蛋白的作用机制尚不清楚。与 T-抗原及 SV40 的早期启动子不同，结构不同的 tax 蛋白和 tat 蛋白都不能直接与LTR 结合。HIV 的 tat 蛋白与一个称为 TAR(反式激活效应)的区域结合，它是一个茎-环结构，位于 LTR 转录产生的 mRNA 的 5′端。有人认为，tat 蛋白与 TAR 区域之间，可能具有类似于 λ 噬菌体中 N 蛋白与 *nut* 位点间那样的转录反终止子功能。虽然这种观点尚不能肯定，但 tat 蛋白和 tax 蛋白有可能通过某种方式，直接或间接地对在病毒 LTR 启动子上形成的转录起始复合体进行修饰。

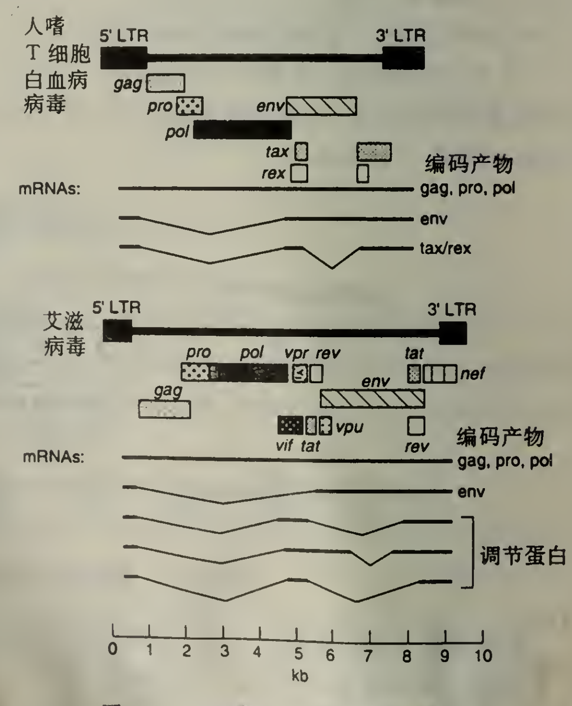


图 11-47　人嗜 T 细胞白血病病毒和艾滋病毒基因组的表达

由于 HTLV tax 蛋白和 HIV tat 蛋白的合成依赖于它们自身活化的启动子，所以，这两种蛋白是由病毒控制的，作用于原病毒 LTR 初始启动子的正调控子，在转录水平上激活所有病毒基因的表达。由于这一过程将导致不能控制的正向反馈作用，并对整合到宿主细胞基因组中的逆转录病毒产生不利影响，所以这一过程是不能单独存在的。因此，HTLV 和 HIV 还分别编码了另外一种蛋白，即 rex 和 rev 蛋白，它们在转录后水平上对基因的表达进行调控，维持粒子结构蛋白和调节蛋白之间的平衡(图 11-48)。

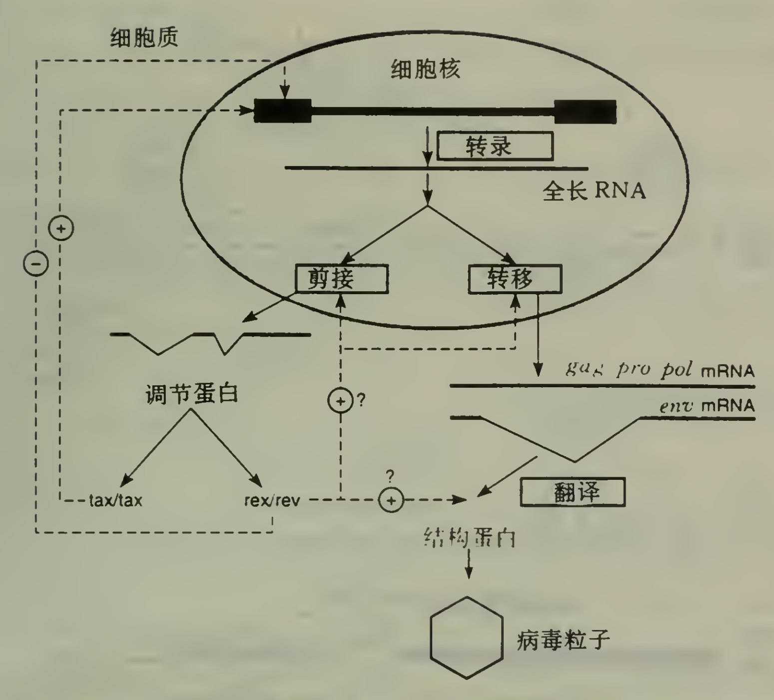


图 11-48　病毒编码蛋白对人嗜 T 细胞白血病病毒和艾滋病毒基因表达的反式调节

五、表达的转录后调控

许多 DNA 病毒可以在细胞核中进行复制，它们编码的 mRNA 在被翻译之前必须借助于寄主细胞的修饰机制进行剪接，去掉中间的部分序列(内含子)。由于 mRNA 在被转移到细胞质进行翻译之前，需要借助于细胞核的功能进行加工，所以这种修饰仅限于一些在细胞核内复制的病毒。另外，还有一些科的病毒则利用寄主细胞的这种能力，将更多的遗传信息"压缩"到它们的基因组中。细小病毒科便是进行这种剪接(splicing)的最好例子。它们的转录可在侵染细胞的细胞质中产生多种经过剪接、带有多腺苷尾序的转录物，使它们 5 kb 的基因组能够编码多种蛋白质。与此相似的还有乳多空病毒科中的猴病毒 40(图 11-44)。相反，由于在疱疹病毒中含有较大的遗传容量，使这些病毒有可能产生几乎不经剪接的单顺反子 mRNA，也就是每一种 mRNA 都是由自身启动子表达的。因此，对这些病毒来说，也就没有必要通过大量的剪接来产生全部所需要的蛋白。

腺病毒基因组的表达是对病毒 mRNA 剪接研究的最好例子之一(图 11-49)。有些类型的腺病毒基因，转录后首先产生核内不均一 RNA(hnRNA)前体，然后再通过各种剪接来进行表达。对于在侵染后立即表达、编码反式作用调控蛋白的早期基因来说尤其是这样。首先，表达的蛋白质是 E1A 和 E1B，是由腺病毒基因组 γ-链靠近左端的一个转录单位编码的(图 11-49)。与上面谈到的 tax 和 tat 蛋白相似，这些蛋白是转录作用的主要反式调控蛋白，同时也与腺病毒感染细胞的转型有关。转录后可以产生 5 种具有多腺苷酸尾序，并经过剪接的 mRNA (13 S，12 S，11 S，10 S 和 9 S)，这些 mRNA 又可编码 5 种相关的 E1A 蛋白(分别为 289，243，217，171 和 55 个氨基酸)(图 11-50)。所有这些蛋白都是由同一个阅读框架翻译而来的，并有

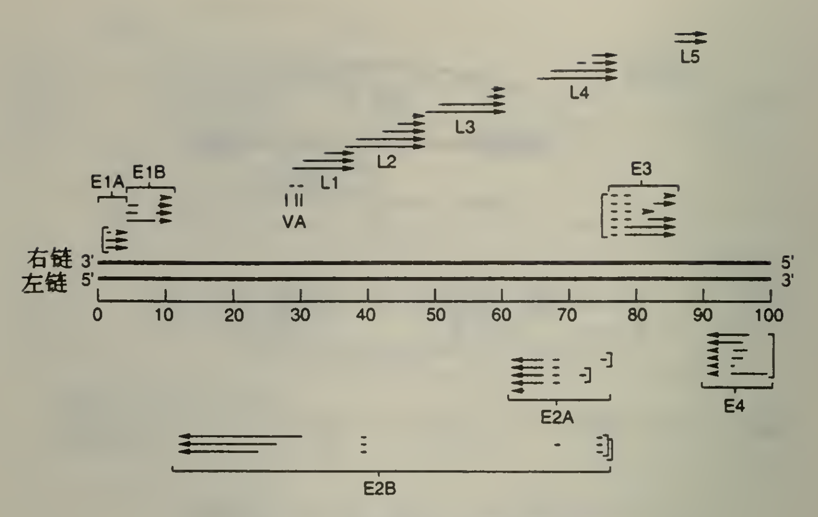


图 11-49　腺病毒基因组的转录

箭头示病毒基因组中外显子的位置，通过剪接产生不同病毒蛋白。

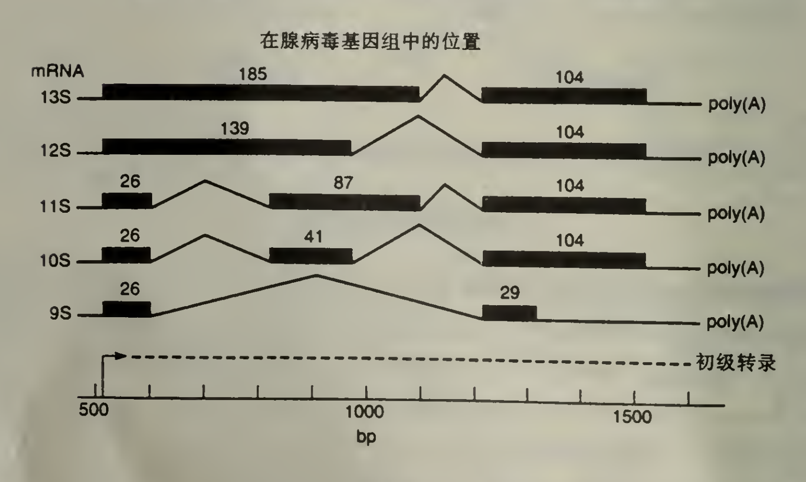


图 11-50　腺病毒 E1A 蛋白的表达

图中数字为每个外显子所编码的氨基酸数目。

着同样的氨基和羧基末端。它们间的差异是对 E1A 转录单位的不同剪接，从而导致了它们功能上的明显差异。289 和 243 个氨基酸的多肽是转录激活子。它们能够激活所有腺病毒早期启动子的转录，并可能还同时具有其它功能。如可以激活许多目前已知的，能够对 RNA 聚合酶Ⅱ进行应答的启动子。这些启动子中并没有明显的共有序列，同时也没有证据表明 E1A 蛋白可以直接与 DNA 结合。现在认为，E1A 蛋白所具有的功能是间接的。它通过对细胞转录因子的化学修饰或与一种 DNA 结合中间蛋白的相互作用结合于启动子上。217，171 和 55 个氨基酸的蛋白不具有转录过程反式激活子的作用，但可能与细胞转型或表达调控有关。

E1A 的合成可以启动 E1B，E2，E3 和 E4 等其它腺病毒早期基因的转录，从而起始一系列转录过程（图 11-49）。在病毒基因组完成复制以后，这一过程最终导致编码结构蛋白的晚期基因进行转录。DNA 病毒的立即早期基因的启动子上游都含有强的增强子单元。在新感染的

细胞中，这种增强子对启动病毒基因的表达是必需的，而合成产生的立即早期蛋白则是起始其它病毒基因表达的转录激活子，如E1A 蛋白便具有这种功能。E1A 的转录是一个平衡的自我调控系统，虽然E1A 蛋白可以反式激活自身的启动子，但是由于它同时可以抑制上游增强子的作用，因此在高浓度下对自身的表达也能进行负调节(图 11-51)。

mRNA 由细胞核到细胞质的转移过程是进行表达调控的另一个阶段，腺病毒中的一些种类仍然是其中研究得最好的例子。在 RNA 聚合酶Ⅲ(它的正常功能是转录产生如 5S rRNA 和 tRNA 等一些小 RNA)的参与下，*VA* 基因在病毒复制的后期阶段，由基因组的 γ-链转录编码两个小的 RNA(图 11-49)，*VA* RNA Ⅰ 和 *VA* RNA Ⅱ。它们都具有高度的二级结构，也不编码任何多肽分子，有着与 tRNA 相似的特点。这两种 RNA 作用方式尚不完全清楚，但它们确实对于提高腺病毒晚期蛋白的合成具有影响。现已经证明，*VA* RNA 与核内小核糖体蛋白(snRNP)的复合物有关，snRNP 是细胞核内 RNA 剪接因子的一部分，但也有证据认为，*VA* RNA Ⅰ 以其它方式选择性起始腺病毒 mRNA 的翻译。病毒侵染细胞后可以刺激产生干扰素，它的作用之一是活化一种细胞中的蛋白激酶，如已知的 DAI 或 p68。这种激酶对细胞和病毒两者的翻译起始都具有抑制作用。*VA* RNA Ⅰ 与这种激酶结合后能够阻碍它的活性，从而解除它对翻译的抑制。也就是 *VA* RNA Ⅰ 可能在细胞 mRNA 的翻译受到抑制的情况下，选择性起始腺病毒 mRNA 的翻译。

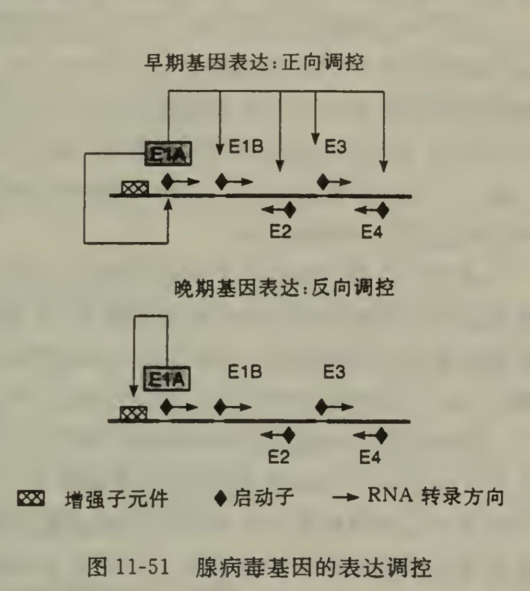


图 11-51　腺病毒基因的表达调控

前面提到的 HTLVrex 蛋白和 HIVrev 蛋白，对病毒特定 mRNA 的翻译也具有选择性。据目前所知，它们能够对病毒基因组的不同表达进行调节，但却不影响细胞 mRNA 的表达。这两种蛋白在氨基酸序列上并不相关，但是它们的作用方式相似，其中 HTLV 的 rex 蛋白在功能上可以代替 HIV 的 rev 蛋白。HIV 和 HTLV 基因组中的某些负调控序列可能会使病毒 mRNA 滞留在感染细胞的核中。这些序列位于内含子内，经过剪接之后，它们将被从 mRNA 中除去，并使mRNA 具有编码 tax，tat 和 rex，rev 蛋白的能力(图 11-47)。通过这些步骤，上述蛋白在侵染之后可以立即得到表达，tax 和 tat 蛋白又可进一步提高由病毒 LTR 引导的转录强度(图 11-48)。一些未经剪接或只经过简单剪接的 mRNA，虽然具有编码 gag，pol 和 env 等蛋白的能力，但只能在细胞中存在足够的 rex 或 rev 蛋白时才能表达。这两种蛋白都是与一种由 mRNA 中的特殊序列形成的二级结构结合。结合以后，它们的作用可能是在核中控制 mRNA由细胞核到细胞质的转移；也可能是在细胞质中通过促进多核糖体的形成或提高它们的活性，来增强 mRNA 的翻译。

由于RNA 的稳定性和二级结构等因素，不同mRNA 的翻译效率有很大差异。其中的一个主要原因可能是翻译起始密码 AUG 前后的核苷酸序列不同，它们是核糖体识别的特殊序列。翻译起始的最佳序列是 GCC(A/G)CCAUGGG，但其中也存在着很大幅度的变异。许多病毒利用这段序列中的差异来对由一种 mRNA 合成的蛋白量进行调控。如在 HTLV 的表达中，tax 和

rex 蛋白是由两个重叠的阅读框架编码的，这两个框架位于同一条经过两次剪接的 2.1 kb 的 mRNA 上(图 11-47)。rex 蛋白的起始密码子 AUG 位于 tax 蛋白起始密码子的上游，但是 AUG 周围的序列对起始翻译过程却不如 tax 蛋白的相应区域合适。因此，在 HTLV 感染的细胞中，尽管两种蛋白由同一条 mRNA 编码，但 rex 蛋白的相对含量明显少于 tax 蛋白。

小 RNA 病毒基因组代表了另外一类翻译起始的调控机制。这些基因组的 5′端含有约占整个基因组的 10%的非编码区(5′-NCR)，并与病毒基因组的复制及包装有关。小 RNA 病毒基因组的翻译方式与大多数细胞 mRNA 不同。在其 RNA 的 5′端没有帽子结构，也不能像其它 mRNA 那样被核糖体识别，但却被一种 VPg 蛋白所修饰。另外，在多蛋白编码区起始位点上游，5′-NCR 中含有多个不能被核糖体识别的 AUG 密码子。实验表明，这类病毒在翻译起始调控中采用了不同的机制。

通过在多蛋白编码区中间插入额外 5′-NCR 信号，构建含有双顺反子的人工突变小 RNA 基因组，并在体外对其 mRNA 进行翻译。结果表明，核糖体的识别对小 RNA 病毒 5′-NCR 内部的某些序列具有特异性，这些序列引导核糖体与 RNA 结合并起始翻译，而不是像细胞 mRNA 那样，由 5′端开始沿着 RNA 按顺序识别。尚未发现在其它病毒中是否也存在这种机制。

另外还有一种称作核糖体移码的机制，被一些病毒用来实现翻译起始调控。其中研究最为深入的是逆转录病毒基因组，冠状病毒和一些芜菁黄花叶病毒组的植物病毒也采用了相似的机制。(这种移码现象目前在某些细菌基因的表达中也已发现，但是除了受到病毒感染以外，还没有在真核细胞中发生的报道。)逆转录病毒基因组转录后至少产生两种 5′具有帽子结构、3′具有多腺苷酸尾序的 mRNA。经过剪接的 mRNA 可以编码包膜蛋白，在 HTLV 和 HIV 等更复杂的逆转录病毒中，还可以编码如 tax 或 tat 和 rex 或 rev 等其它蛋白(图 11-47)。一种长的未经剪接的转录物可编码 *gag*，*pro* 和 *pol* 基因，同时也作为包被在病毒粒子中的基因组 RNA。在不同的逆转录病毒中，这 3 个基因的排列是不一样的。在某些情况下(如 HTLV)，它们位于 3 个不同读码框架中；而在其它的病毒中(如 HIV)，蛋白酶(*pro*)基因是由 *pol* 基因的 5′端延伸构成的(图 11-52)。在后一种情况下，蛋白酶和聚合酶(即反转录酶)是作为一种多蛋白来表达的，经过自身切割后产生成熟蛋白，过程与小 RNA 病毒多蛋白的切割相似。

在这 3 个基因之间的边界上，存在一段通常由重复核苷酸构成的特殊序列，如UUU-AAAC(图 11-53)，大多数核糖体都可以顺利地对它进行翻译。但是，有一部分核糖体在对这种序列进行翻译时，却需要先向后滑动一个核苷酸，然后才能以不同的阅读框架(即－1 位)继续翻译过程。因此，UUUAAAC 序列可称为“滑动序列”(slippery sequence)，移码的结果是翻译得到一种由不同阅读框架而来的多蛋白。同时，这种机制也使病毒能够对所产生的蛋白比例进行调控。由于仅有部分核糖体在滑动序列上发生移码，所以从 mRNA 阅读框架 5′端到 3′端的翻译存在着一种梯度。

然而，滑动序列本身仅能导致较低频率的移码，这对于产生病毒所需的蛋白酶和反转录酶的量来说是不够的。因此，有另外一类序列在这一系统中，通过增加移码的发生频率，发挥着进一步的调控作用。离滑动序列不远的下游，有一个反向重复区，它使 mRNA 中形成一个茎-环(stem-loop)结构(图 11-53)。在稍远一些的位置上，还存在另外一段与环状结构核苷酸互补的序列，它使 RNA 的这两个区域间发生碱基配对。所形成的结构也就是通常所说的 RNA 假结节(pseudoknot)。mRNA 中的这种二级结构，使核糖体的翻译在上游的滑动序列中放慢或中止，提高了移码的发生频率，也就增加了下游框架编码蛋白的相对量。

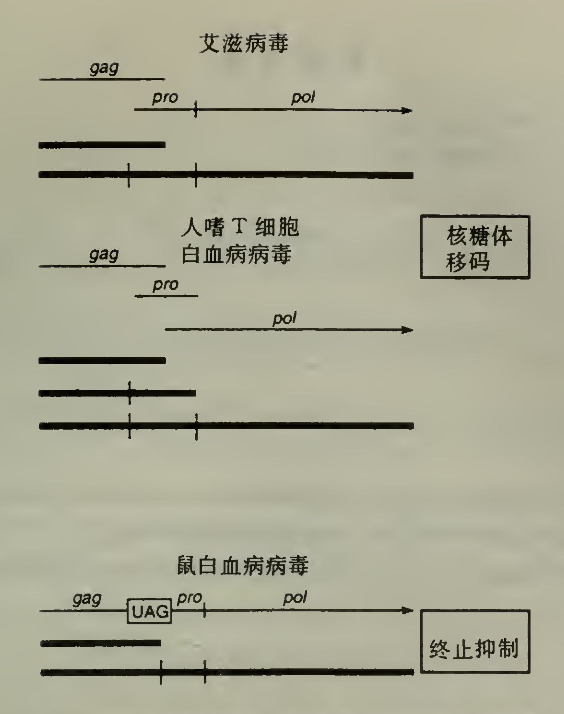


图 11-52　逆转录病毒中的核糖体移码和终止抑制

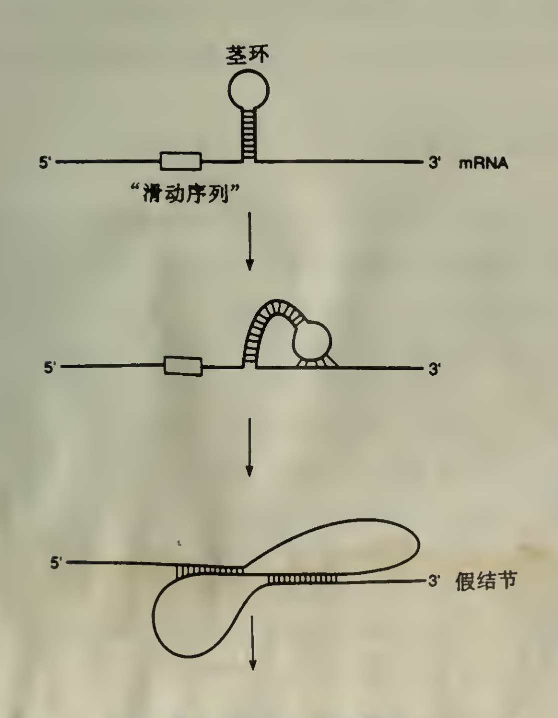


图 11-53　RNA 假结节的形成和核糖体移码的机制

翻译调控的最后一种途径是终止抑制。这种机制在许多方面与移码相似，它可以在一条mRNA上，通过不同的阅读框架表达产生多种蛋白。如逆转录病毒中的小鼠白血病病毒(MLV)，它的 *pro* 基因与 *gag* 基因之间不存在"滑动序列"和假结节，而是被 UAG 终止密码子分隔开来(图 11-52)。在大多数情况下，MLV mRNA的翻译将在此序列上终止，产生gag 蛋白。但是在有些情况下，UAG 终止密码子会受到抑制，继续翻译产生一个gag-pro-pol 多蛋白，然后再切割产生成熟蛋白。gag，pro 及pol 蛋白的相互比例，受核糖体对于UAG 终止密码子的通读频率调控。

参考文献

1. 侯云德. 分子病毒学. 北京:学苑出版社,1990.
2. 裘维藩主编. 植物病毒学. 北京:科学出版社,1985.
3. Butler, P. J. G. and Klug, A. The Assembly of a Virus. Scientific American. 1978. 239:62～69.
4. Cann, A. J. Principles of Molecular Virology. Academic Press, 1993.
5. Darnell, J., Lodish H. and Baltimore D. Molecular Cell Biology (2nd ed.). Scientific American Books. 1990.
6. Dolia, V. V. and Carrington, J. C. Evolution of Positive-strand RNA Viruses. Seminars in Virology. 1992. 3:315～326.
7. Fraenkel-Conrat, H., Kimball, P. C. et. al. Unenveloped Plus-strand RNA Viruses, in Virology. Prentice Hall, 1988. p. 1～59.
8. Harrison, S. Principles of Virus Structure, in Virology. B. N. Raven Press, 1985. p. 27～44.
9. Matthews, R. E. F. Plant Virology (3rd. ed.). Academic Press, 1991.
10. Waterson, A. P. and Wilkinson, L. An Introduction to the History of Virology. Cambridge University Press, 1978.

（于嘉林）

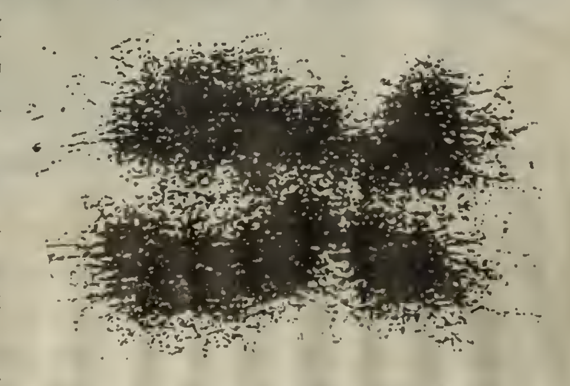
第十二章　真核染色体和基因组的结构

真核染色体比原核的大，并且具有高度的结构性组织。酵母、果蝇和人类基因组所含的DNA分别约为大肠杆菌基因组的4，40和1 000倍。这样大的基因组使得真核生物具有了原核生物所没有的潜力，同时也使真核基因的表达遇到额外的挑战。

本章将重点讨论以下几个问题：①真核的很长的DNA（人类总DNA的线型长度接近1 m）是怎样压积在细胞核中的？和原核的裸露的DNA不同，真核DNA是与一类叫做组蛋白的碱性蛋白质结合在一起的，完整的这种DNA-蛋白质复合体叫做染色质。在DNA的压积中组蛋白起着关键性作用。DNA缠绕在组蛋白上形成核小体，许多核小体由DNA连在一起形同串珠。它再经折叠压积而形成染色体。真核染色体的这种结构对基因的表达有着重要的影响；②在真核染色体中基因是怎样安排的？高等真核生物基因组的一个突出特点是它含有大量的重复序列，其中绝大多数是不编码蛋白质或RNA的，而且在编码蛋白质的基因中绝大多数是断裂的，即在编码片段（外显子）之间含有非编码的片段（内含子）。因此，在高等真核生物的基因组中只有很小一部分是编码蛋白质的。此外，在真核生物的线粒体和叶绿体中也含有基因组。这些细胞器中的基因组在结构上有着与核基因组不同的特色。最后，由于真核基因是断裂的，而且在转录时内含子与外显子一起转录，因而转录产生的RNA必须经过加工，将内含子剪接掉，才能成为成熟的mRNA，作为蛋白质合成的模板。我们将讨论这种剪接的机制，以及由于对剪接机制的研究，导致了核酶的发现。

第一节　真核染色体的结构

真核染色体是由染色质（chromatin）构成的，染色质是由DNA、RNA和蛋白质形成的复合体。染色体是动态的物体，其外观随细胞周期的不同阶段发生明显的改变。仅当细胞分裂（细胞周期的M期）时，每个染色体才呈现出大家所熟知的凝聚型（图12-1）。在分裂间期（细胞周期的其余阶段），染色体DNA发生转录和复制，此时大多数细胞的染色体变为高度弥散，以致于无法区分每个染色体。长期以来细胞学家即已辨认出这种弥散的染色质有两种型：一种密度较低，称为常染色质（euchromatin）；另一种密度较高，称为异染色质（heterochromatin）。这两种染色质的区别是：常染色质被表达，异染色质不被表达。事实是，真核染色体的结构对其基因的表达有着明显的影响，而原核的影响则不明显。因此，在研究真核基因的表达时必须要研究其染色体的结构。


图12-1　人类M期染色体的电镜图

真核DNA是怎样组装在染色体中的？这个重要问题曾经困惑了人们许多年。经过多年

的研究，现在已知一个真核染色体中含有一个完整的DNA分子，并且此DNA分子是线型的和不具分支的。其分子量很大。例如，果蝇最大染色体中DNA的分子量为43×10^6 kD，相当于76×10^6 bp，约为大肠杆菌的20倍(表12-1)。

在人类细胞的6个染色体中，每个染色体所含的DNA为4 800万～24 000万bp，它们极可能是完整的，因而其线型长度在1.6～8.2 cm之间。然而，在细胞分裂期，即染色体处于最凝聚状态时，每个染色体的长度却仅为1.3～10 μm之间。可见，染色体DNA的堆积比(packing ratio)(指其线型长度与其容器长度的比值)大于8 000。染色质中的DNA是怎样达到如此高度凝聚的？结构研究揭示，这是通过3个水平的折叠做到的，下面逐步讨论之。

表12-1 一些原核和真核基因组中DNA的含量

生物	碱基对数(bp)	DNA长度(mm)	染色体数
大肠杆菌	4×10^6	1.4	1
酵母(*S. cerevisiae*)	1.4×10^7	4.6	16
果蝇(*D. melanogaster*)	1.7×10^8	56	4
人	3.9×10^9	990	23

注：为单倍体基因组数值。

一、组 蛋 白

和原核不同，真核染色体中的DNA不是裸露的，而是与蛋白质紧密地结合在一起。与真核DNA相结合的蛋白质主要是一类小的碱性蛋白质，称为组蛋白(histone)。组蛋白约占真核染色体总量的一半，其余一半为DNA。此核蛋白的染色物质称为染色质。用盐或稀酸处理染色质，可使组蛋白与DNA解离开来，再用离子交换柱层析分级分离组蛋白，共得到5个组分。这5种组蛋白分别命名为H1，H2A，H2B，H3和H4。它们的分子量在约11 kD～21 kD之间(表12-2)。组蛋白的一个突出特点是含有很多带正电荷的氨基酸残基，约4个残基中就有1个是赖氨酸或精氨酸。因此，组蛋白是以离子键与DNA中带负电荷的磷酸基相结合的，所以用0.5mol/L的NaCl溶液即可把组蛋白由染色质中抽提出来，因为这样浓度的盐溶液即足以破坏其静电作用。

表12-2 小牛胸腺组蛋白

组蛋白	残基数	分子量(kD)	%精	%赖
H1	215	23.0	1	29
H2A	129	14.0	9	11
H2B	125	13.8	6	16
H3	135	15.3	13	10
H4	102	11.3	14	11

组蛋白在进化中是保守的。氨基酸顺序分析表明，豌豆秧苗的H4和小牛胸腺的H4的氨基酸顺序，在它们的102个残基中仅有两个差异。而且这两个位点的改变颇小，即一个位点是由异亮氨酸变为缬氨酸；另一个位点是由精氨酸变为赖氨酸。可见，自从植物和动物趋异后的1.2×10^9年中H4的氨基酸顺序几乎没有改变。同样，在经过这样长的进化期间H3的改变也很小，豌豆秧苗H3的氨基酸顺序和小牛胸腺的仅有4处不同。将在进化过程中这些组蛋白的改变速度与其它蛋白质的改变速度进行比较是有意义的。一个有用的指数叫做单位进化周期(unit evolutionary period)，它指的是在两个进化系(evolutionary lines)趋异后顺序改变1%的时间间隔。对于H3来说此间隔为3亿年，H4为6亿年，比迄今已研究过的其它蛋白质长很多。例如，细胞色素c的单位进化周期为2×10^7年，血红蛋白为6×10^6年，血纤肽(fibrinopeptides)为1×10^6

年。H3 和 H4(还有 H2A 和 H2B)结构的高度保守,说明在真核生物进化的初期即已确定它们的作用至关重要,而它们的结构也已完全适合它们的功能,但不能忍受改变。至于 H1 则比其它组蛋白更可改变,下面我们将要谈到这是由于它的功能与其它组蛋白不同之故。

组蛋白在翻译后是受到修饰的,其中包括特异精、组、赖、丝和苏氨酸残基的甲基化、乙酰基化和磷酸化(图 12-2)。这些修饰中有许多是可逆的。所有修饰都降低组蛋白的正电荷,因而显著改变组蛋白-DNA 之间的相互作用。虽然组蛋白在进化中是高度保守的,但它们修饰的程度却随品种、组织以及细胞周期阶段的不同而差别很大。一个特别引人注目的修饰是 10%的 H2A 联结上了泛素(ubiquitin),即它们的赖(119)的 ε-氨基与泛素的羧基之间以肽键联结在一起。虽然已知这种泛素化(ubiquitination)的作用是给胞液中需要降解的蛋白质打上标签,从而促使这些蛋白质降解,然而还不知道 H2A 的泛素化是否也是起这种作用。最令人感兴趣的问题是,这种泛素化以及组蛋白的其它方式的修饰是否是以某种方式调控着真核基因的表达。

Ac— Ser—Gly—Arg—Gly—Lys—Gly—Gly—Lys—Gly—Leu— 10
Gly—Lys—Gly—Gly—Ala—Lys—Arg—His—Arg—Lys— 20
Val—Leu—Arg—Asp—Asn—Ile—Gln—Gly—Ile—Thr— 30
Lys—Pro—Ala—Ile—Arg—Arg—Leu—Ala—Arg—Arg— 40
Gly—Gly—Val—Lys—Arg—Ile—Ser—Gly—Leu—Ile— 50
Tyr—Glu—Glu—Thr—Arg—Gly—Val—Leu—Lys—Val— 60
Phe—Leu—Glu—Asn—Val—Ile—Arg—Asp—Ala—Val— 70
Thr—Tyr—Thr—Glu—His—Ala—Lys—Arg—Lys—Thr— 80
Val—Thr—Ala—Met—Asp—Val—Val—Tyr—Ala—Leu— 90
Lys—Arg—Gln—Gly—Arg—Thr—Leu—Tyr—Gly—Phe— 100
Gly—Gly 102

图 12-2 小牛胸腺组蛋白 H4 的氨基酸顺序

在此 102 个残基的蛋白质中精和赖残基占了 25%。豌豆秧苗的 H4 与小牛胸腺的 H4 有两个残基不同:缬(60)⟶异亮和赖(77)⟶精。下面划线的是翻译后修饰的残基:丝(1)总是 N-乙酰基化的,它也可发生 o-磷酸化,赖(5,8,12 和 16)可 N-乙酰基化,赖(20)可被 1 或 2N-甲基化。

许多真核生物具有遗传上不同的 H1,H2A,H2B 和 H3 的亚类,它们是在胚胎发生和某些类型细胞发育的特异阶段合成或不合成。对于 H2A,H2B 和 H3 来说,这些亚类的序列改变仅为极少数氨基酸残基,但 H1 的改变则比较多。组蛋白合成的开关似乎与细胞分化有关,但这种关系的本质尚不清楚。

二、核小体:染色质结构的第一层次

组蛋白是怎样与 DNA 结合起来以形成染色质纤维的? 1974 年,Roger Kornberg 综合多方面的实验证据指出,染色质是由重复单位构成的,每个单位由约 200 bp 的 DNA 和 H2A,H2B,H3,H4 各 2 分子所组成。现在,把这些重复单位叫做核小体(nucleosome),它是染色质结构的第一层次。在核小体中,大部分 DNA 缠绕在组蛋白构成的核心(core)的外面。DNA 的其余部分将相邻的核小体联结起来,此部分 DNA 称为联结物(linker),它使染色质纤维具有柔性。可见,一条染色质纤维是条由联结起来的核小体构成的易弯曲的链,很像一条串珠。染色质的这一结构模型具有广泛的实验证据,主要如下:

(1)染色质所含 H2A,H2B,H3 和 H4 的分子数目大致相等,H1 的分子数不超过它们的一半。

(2)电镜观察(图 12-3)染色质，发现它由直径约 10 nm 的颗粒所组成，这些颗粒由裸露的 DNA 细线联结在一起。可见，电镜观察直接提供了染色质结构的概念，即它是用易弯曲区隔开的一串大致为球形的颗粒。

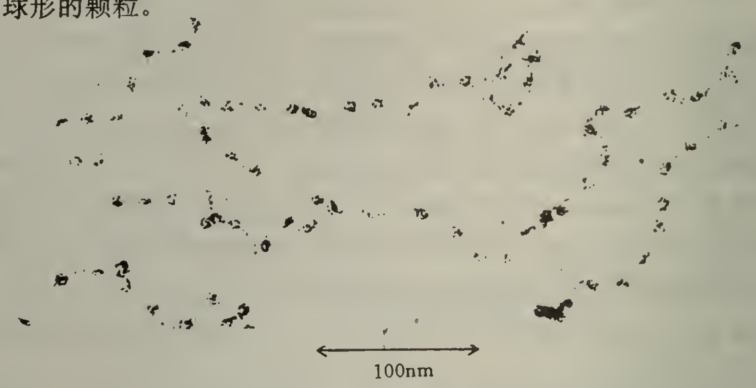


图 12-3 染色质的电镜图

串珠样颗粒的直径约为 10 nm。引自 Ada Olins 和 Donald Olins.

(3)X 射线衍射研究得出了与电镜观察相同的结论。中子衍射的研究表明 DNA 位于核小体的外侧。

(4)核酸酶消化实验。用胰脱氧核糖核酸酶 1(DNase 1)或小球菌核酸酶在溶液中消化游离 DNA 时可裂解其任意的磷酸二酯键。与之相反，染色质中的 DNA 被保护了，仅在少数位点处能被酶裂解。酶解染色质的电泳图谱的突出特征是它的简单性：其图谱(图 12-4)由一系列不同大小的带所构成。这些片段中的 DNA 含量是约为 200 bp 的基本单位的倍数。电镜观察表明，染色质一个片段中球形颗粒的数目与其 DNA 为 200 bp 单位的倍数相同(图 12-5)。例如，含有 600 bp DNA 的片段是由 3 个直径为 10 nm 的颗粒组成的。可见，在电镜下观察到的一个珠子相当于用核酸酶消化确定的一个核小体。



图 12-4 用小球菌核酸酶部分消化染色质后，消化产物的凝胶电泳图谱

A. 为未经分级分离的消化产物。蔗糖梯度离心分离后产生：B. 单体；C. 二聚体；D. 三聚体；E. 四聚体。

(5)再造。将组蛋白加到腺病毒或 SV40 的 DNA 中可在体外形成染色质样纤维。在这些再造体系中，与一个核小体相结合的 DNA 极近于 200 bp。而且，形成核小体所需 H2A，H2B，H3 和 H4 的分子数目是相等的。在再造混合物中缺少这 4 种组蛋白中的任何一种都不能形成这种特点的珠子。但 H1 是不需要的，这与 H1 不整合在核小体中的发现相一致。

(一)DNA 盘绕在一个组蛋白八聚体上形成核小体的核心颗粒

如上所述，用小球菌核酸酶降解染色质开始生成的是单个核小体，它还含有组蛋白 H1。进一步消化时其 DNA 被剪短并释出 H1。此产物称为核小体的核心颗粒，它由 146 bp 的 DNA 和一个组蛋白的八聚体(H2A，H2B，H3 和 H4 各 2 分子)所组成。在所有真核生物中的核小体核心几乎都是相同的。原来联结在核小体上并被进一步剪去的 DNA 部分称为联结物，它的长度虽然一般为

55 bp 左右，但在不同生物和同一生物的不同组织中的变化却很大，在 8～114 bp 之间，从而使不同生物和同一生物的不同细胞类型的核小体 DNA 含量在约 160～240 bp 的范围内变动(表 12-3)。

表 12-3　核小体的 DNA 含量

细胞类型	碱基对数(bp)
酵母	165
Hela	183
大鼠骨髓	192
大鼠肝	196
大鼠肾	196
鸡输卵管	196
鸡红细胞	207
海胆原肠胚	216
海胆精子	241

Gerard Bunick 和 Edward Uberbacher 用 X 射线分析了核小体核心颗粒的结构(图 12-6)，分辨率在 0.8 nm 左右。在此低分辨率的情况下，仅能辨别出来诸如 α 螺旋之类的分子的粗略情况。核心颗粒是个双折叠对称的盘子形，厚 7 nm，直径 11 nm。$(H3)_2(H4)_2$ 四聚体构成组蛋白八聚体的核心，盘子的顶部和底部各有一个 H2A · H2B 二聚体。B-DNA 以左手超螺旋形式缠绕组蛋白八聚体 1.8 圈，螺距为 2.8 nm。然而，DNA 并不遵循平稳的超螺旋途径，而是在有些局部做相当急剧的弯曲，以致其大沟和小沟的宽度发生大的改变。蛋白质-DNA 的相互作用发生在 DNA 超螺旋的内侧，没有蛋白质缠绕 DNA 的现象，蛋白质也不由 DNA 超螺旋的圈与圈之间凸出。

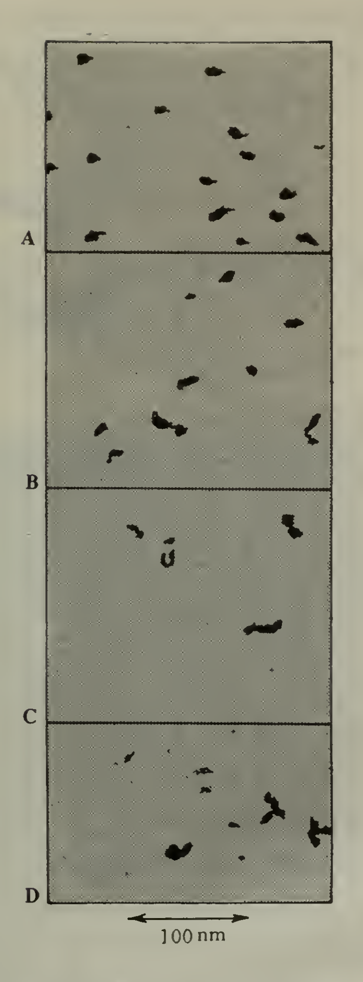


图 12-5　图 12-4 中的单体(A)，二聚体(B)，三聚体(C)和四聚体(D)的电镜观察

(二)组蛋白 H1 把核小体“封锁”起来

在用小球菌核酸酶消化核小体的约 200 bp DNA 时首先降解为 166 bp。然后经过暂时停顿之后，H1 释出，DNA 进一步降解为 146 bp。核心颗粒的双折叠对称(twofold symmetry)结构提示，原来长度为 166 bp-的 DNA 是由其两端各移去约 10 bp 而变为 146 bp 的。由于核心颗粒的 146 bp DNA 形成 1.8 超螺旋圈，因而 166 bp 中产物应能形成完整的 2 超螺旋圈，这样将使它的两个末端尽可能地靠近到一起。因此，Klug 提出，H1 是在核小体 DNA 的中部段与 DNA 进入和离开核心颗粒的片段之间形成的空洞处联结在核小体 DNA 上的(图 12-7)。支持此 模型的实验证据是：染色质纤维含有H1，DNA是在同一侧进入和离开核小体，但在没有H1 的染色质中，则 DNA 的进、出位置是更为随机分布的，并且趋于发生在核小体的对侧。此模型还提示，联结物 DNA 的长度是由联结在其上的组蛋白 H1 的亚类决定的。

三、30 nm 纤丝：染色质结构的第二层次

166 bp 核小体DNA 的堆积比约为 10(其 56 nm 线型长度弯曲成一个约 5.6 nm 高的超螺

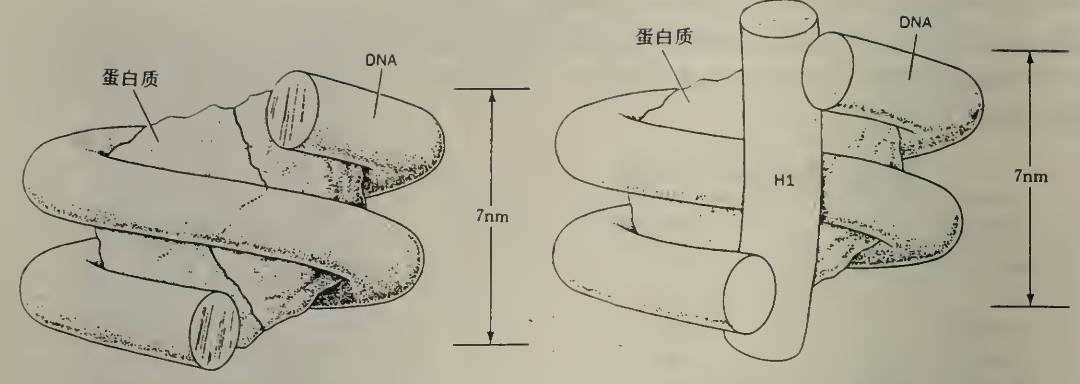


图 12-6　核小体核心颗粒模式图

图 12-7　设想组蛋白 H1 是联结在 166 bp 核小体上的。DNA 的两个完整超螺旋圈使 H1 能够与 DNA 的两端及其中部联结起来

旋）。显然，在低离子强度下形成 10 nm 核小体纤丝（filament）仅代表染色体 DNA 堆积的第一层次。只有在生理的离子强度下，染色体结构的下一层次才被明显看出来。

随着盐浓度的提高，含有 H1 的核小体纤丝开始折叠成曲曲折折的构象（电镜下可观察到）。此外，观察提示，核小体通过它们的H1分子间的接触而相互作用。然后，当盐浓度接近生理范围时，染色质形成一条 30 nm 厚的纤丝（图 12-8），其中的核小体是可以看到的。Klug 认为，30 nm 纤丝是由 10 nm 核小体纤丝卷绕成为圆筒形线圈而形成的，此线圈的每一圈约为 6 个核小体，螺距为 11 nm（核小体直径，图 12-9）。此圆筒形线圈是由 H1 分子稳定的。每个 H1 分子由两部分组成：一部分是保守的球形核

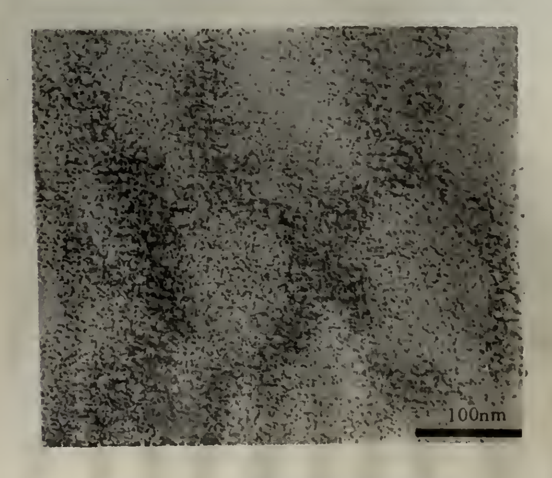


图 12-8　30 nm 染色质纤丝的电镜图
注意纤丝为 2～3 核小体宽。线段表示 100 nm 长。引自 Jerome B. Rattner.

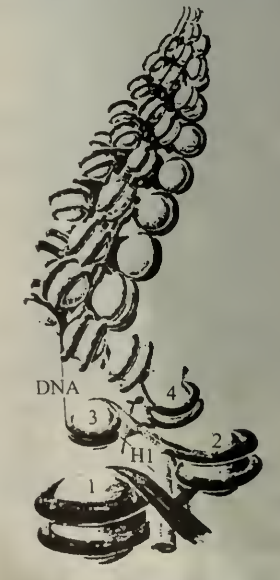


图12-9　30 nm 染色质纤丝的一种模型表示出随着盐浓度的升高纤丝的可能形成过程
之字形的核小体(1,2,3,4)靠近在一起形成一个圆筒形线圈，每圈约 6 个核小体。许多 H1 分子稳定此结构，设想它们是沿着圆筒形线圈的中心形成螺旋形多聚体。

心，它是联结核小体的；另一部分是比较可变的延伸出去的 N 端和 C 端臂，设想它们是与相邻的核小体相接触。此模型与 30 nm 纤丝的 X 射线衍射图谱相符合，其堆积比约为 40(6 个核小体，每个约 200 bp DNA，总高 11 nm)。然而，需要注意的是，对于 30 nm 纤丝来说，还提出了几种其它的模型，这些模型看来似乎也是合理的。

四、辐射的环(radial loops)：染色质结构的第三层次

组蛋白被剥离的(细胞分裂)中期的染色体表现出一个中心纤维状的蛋白质"脚手架"，其周围环绕着广泛的 DNA 晕圈(halo)(图 12-10 a)。在



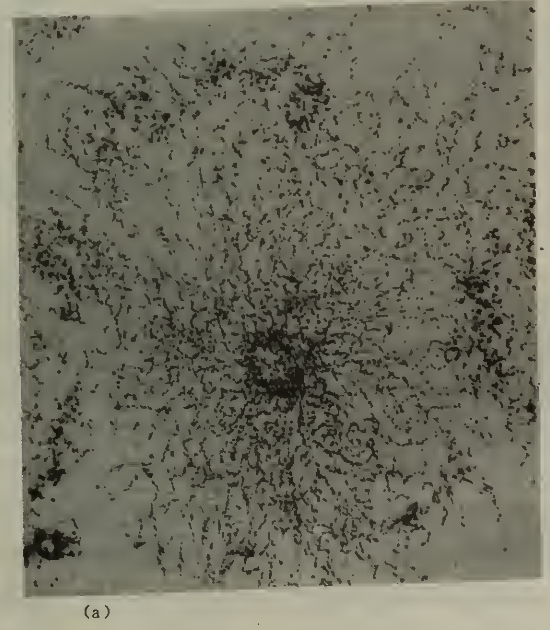
(a)



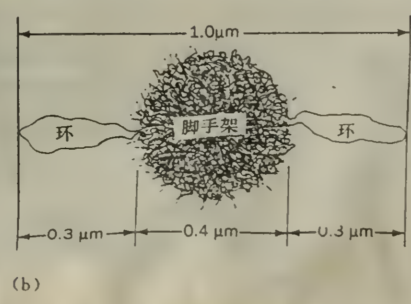
(b)

图 12-10 组蛋白被剥离的人类中期染色质电镜图

(a)中心蛋白质基质(脚手架)，作为周围 DNA 的支撑点；(b)更高倍的放大，可以看到 DNA 以环形连到脚手架上。引自 Ulrich Laemmli.



(a)



(b)

图12-11 (a)人类中期染色体横断面电镜图，注意大量的染色质纤维辐射样的由中心脚手架射出；(b)解释(a)图的模式图

指明设想的 0.3 μm 长的辐射环是怎样联结在 0.4 μm 宽的脚手架上，以形成直径为 1.0 μm 的中期染色体的。引自 Ulrich Laemmli.

电镜下可以看到DNA的股线是成环的，它几乎是由同一位点进出脚手架(图12-10 b)。这些环的大多数的长度在15～30 μm(相当于45～90 kb)范围内，因而当它们堆积成30 nm纤丝时，其长度应为约0.6 μm。电镜观察染色体的横断面，如图12-11 a所示，有力地提示中期染色体的染色质纤维是辐射状安置的。如果看到的环相当于这些辐射纤维，它们的每一个将使染色质的直径加大0.3 μm(一条纤维必须回到其原点而形成环)。加上脚手架的宽度为0.4 μm，此模型指明中期染色体的直径为1.0 μm，这与观察到的是一致的(图12-11 b)。一条典型的人类染色体含有约14×10^7 bp，因而它应有2 000个左右这种约70 kb的辐射环。这样一条直径为0.4 μm，长6.0 μm的染色体脚手架，其面积足以联结2 000个左右的辐射环。因此，辐射环模型是符合观察到的中期染色体的堆积比的。

关于30 nm纤丝是怎样组织起来以形成辐射环的情况几乎什么都不清楚，也不知道中期染色体与更为松散的细胞分裂间期(inter-phase)染色体之间是怎样互变的。然而，可以肯定的是，几百甚至是几千种不同的非组蛋白蛋白质(它们约占染色体蛋白质总量的10%)必定参与这些过程。而且，有证据表明辐射环是染色体的转录单位。

第二节　真核基因组的结构

真核生物的特征是它的基因组的主要部分存在于核中，并用核膜将其保护起来而不暴露在胞浆中。另外有一小部分存在于细胞器中。在我们探讨真核生物的基因组结构时，首先要提出的一个问题是，在高等真核生物中其不同细胞所含的基因组是否相同?高等生物具有很多种不同的细胞，它们不仅在外观上不同，而且合成的蛋白质也不相同。例如，胰脏的腺细胞合成大量的消化酶类，但不合成胰岛素，而与之邻近的胰β细胞却产生大量的胰岛素，但不产生消化酶类。很显然，不同类型的细胞所表达的基因是不同的。然而一个多细胞生物的大多数(不是全部)体细胞所含的遗传信息是相同的，因为它们都是从一个受精卵传下来的，此现象被说成是全能性(totipotency)。例如，John Gurdon曾经证明了这一点。他用蝌蚪肠细胞的细胞核取代青蛙受精卵的细胞核，并用这种受精卵培育出来了正常的成年青蛙(关于细胞是怎样分化的将在第十三章中讨论)。

然而，不同生物核中的DNA总量是差别很大的，组成核基因组的序列类型也可以是很不相同的，不同生物的染色体数目差别也很大，细胞器中的基因大小也不一致。当然，不同真核生物的基因组之间也会有些共同的特点，在这一节中我们就是要讨论真核生物的核基因组结构的某些共同特点(关于细胞器基因组将在第三节中讨论)。但是，必须注意到真核生物界是非常之大的，而我们现在知道的只是少数几种生物的基因组结构。因此，我们在下面讨论真核DNA的特点时，只能说是在已知的内容中挑选出一些有代表性的特点加以分析，而不能确定真核基因组的所有结构特点。

一、*C*-值悖理

在讨论真核生物基因组的结构时，我们首先看看真核基因组的大小。人们早已知道真核细胞DNA的含量比原核的大很多，例如，哺乳动物细胞的DNA含量约为大肠杆菌的800倍。在真核生物中，每种生物的单倍体基因组的DNA总量是恒定的，称之为*C*-值(*C*-value)，*C*-值是每种生物的一个特性。不同物种的*C*-值差别很大，图12-12列举不同门(phyla)生物*C*-值的范围。

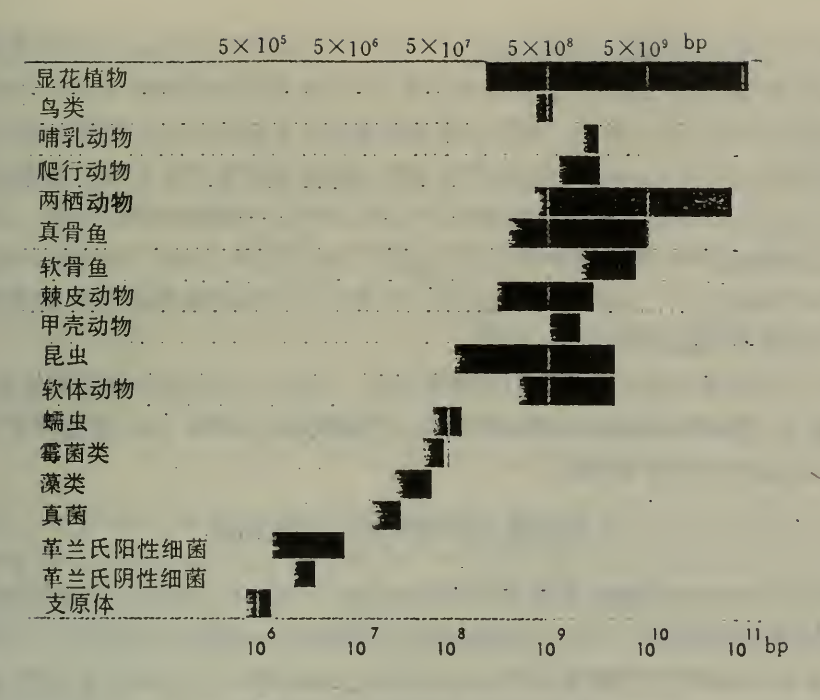


图 12-12　不同门生物的 C-值范围

人们有理由设想，一个生物的形态学复杂性应该与其 C-值的大小大致相关，因为归根结底，一个生物的形态学复杂性必然是它的基因复杂性的反映。从图中可以看出，低等生物的情况大致符合这种设想。例如，啤酒酵母(*S. cerevisiae*)的基因组约为 1.3×10^7 bp，比细菌(大肠杆菌为 4.2×10^6)大约 3 倍，它是单细胞真核生物。而粘菌(slime mold，*D. discoideum* 的基因组为 5.4×10^7)的基因组比酵母的更大，从而使它既能以单细胞形式生长又能以多细胞形式生长。而首先出现的是完全多细胞生物——线虫(nematode worm，*C. elegans* 的 DNA 含量为 8×10^7)的基因组又比粘菌的大。然而，在生物进化为更加复杂时情况就并非如此了。例如，有些两栖类的基因组与人的同样大，难道它们的基因的复杂性真的和人一样吗?而且许多生物的 C-值是出乎预料的大，例如，肺鱼(lungfish)的比哺乳动物的大 10～15 倍，这就更难以理解。再者，从图 12-12 中可以看出，有些门，如鸟类、爬行动物和哺乳动物的 C-值范围是比较窄的，最大的约为最小的两倍；然而在另些门，如昆虫、两栖动物和植物，其 C-值的变动范围很大。一种普通家蝇的基因组为 8.6×10^8，一种普通果蝇的基因组为 1.4×10^8，家蝇比果蝇的约大 6 倍。人们怀疑家蝇的形态学是否真比果蝇复杂 6 倍。这种形态学的复杂程度与 C-值大小的不一致称为 C-值悖理(C-value paradox)。它使人们怀疑，在比较大的基因组中的“额外”DNA 是否有功能? 如果没有功能，为什么还一代一代地传了下来?

引起这种疑问的另一个问题是对基因数目的估测。据认为，大肠杆菌的 4×10^6 bp DNA 约编码 3 000 种基因产物(其中一半左右已被鉴定出来了)。人类单倍体基因组为 29×10^8 bp，比大肠杆菌的大 700 多倍，按大肠杆菌的标准估算，应该有上百万个基因。然而事实却远非如此。尽管到目前为止还无法测知人类(或哺乳动物)基因组中确实含有多少个编码蛋白质的基因，但可以肯定比上述估算的数字要少很多。当前，有多种估测基因数目的方法，其中之一是根据不同细胞中 mRNA 的数目来估算每种细胞中表达产生蛋白质的基因数目。根据这种估算，

哺乳动物的每种表型的细胞表达的基因约为 1×10^4 个。由于有些基因是在所有类型的细胞中都表达的，即这些基因的功能为所有细胞所必需，因而常把它们叫做持家基因(housekeeping gene)或组成型(constitutive)基因。而另一些基因则是仅在某种类型的细胞中表达的，有时把它们叫做奢侈基因(luxury gene)。这样，整个哺乳动物的基因数要多于每种细胞的表达数，估计应该有 1×10^5 个(不同作者估算的数目有所不同)，即约为大肠杆菌的 30 倍。由此可见，在人类以及其它真核生物的基因组中存在着大量的不被表达的 DNA(人类的约占总 DNA 的 98%)，它们的功能是什么？诚然，真核基因表达的调控比原核的要复杂得多，难道这些不被表达的 DNA 的功能就是调控基因的表达吗？

当前对于上述这些问题是无法给以圆满解答的。然而，从后面的讨论中将会看到，我们已经获得了大量关于真核染色体基因结构的知识，并将继续深入研究下去。这些知识对于找到上述问题的答案无疑是有重大帮助的。

二、复性动力学指示 DNA 的复杂度

DNA 复性(renature)的速度能够指明该DNA 的单一序列(unique sequence)的长度。Roy Britten 及其同事在研究热变性 DNA 的复性动力学时发现，与原核生物的 DNA 不同，真核生物 DNA 中含有许多重复的碱基序列。在实验中，他们将 DNA 剪切成小的片段(300～10 000 bp)，然后将含有这些小片段 DNA 的溶液加热至温度高于 DNA 的解链温度(melting temperature，T_m)，使 DNA 变性。再将此含有单链 DNA 的溶液温度降至约 25℃(低于 T_m)，这是互补的单链DNA 重新结合起来成为双螺旋DNA 的最适温度。有多种方法测定其再结合的动力学，其中比较好的一种是利用羟磷灰石(磷酸钙)柱层析。此法的原理是，由于羟磷灰石仅与双链 DNA 相结合，不与单链 DNA 结合，因此，当含有二者的混合溶液流经羟磷灰石柱时，双链 DNA 结合在柱上，单链 DNA 则流出，从而可将二者分开，并能计算出在给定时间内变性的 DNA 中有多少复性了。

用大肠杆菌或 T4 噬菌体 DNA 所做的实验结果表明，它们的复性动力学符合双分子反应

$$\mathrm{S}+\mathrm{S}'\xrightarrow{k}\mathrm{D}$$

规律，这是事前所预期的，因为此反应是由互补的两条单链(式中以 S 和 S′代表)结合成一条双链(式中以 D 表示)。式中的 k 为结合反应速度常数。对于这样的反应来说，其单链分子随时间的延长而下降的分数(f)可按下列公式计算：

$$f=\frac{1}{1+k\,C_o t}$$

式中 C_o 为 DNA 的总浓度(以每升中核苷酸的摩尔数表示)，t 为时间(s)。在给定的 DNA 和实验条件(例如离子强度、温度、DNA 片段的大小)下；f 仅随 $C_o t$(DNA 浓度和时间的乘积)的改变而改变。描述复性动力学的方便方法是以 f 对 $C_o t$ 的对数做图，这种 $C_o t$ 曲线为 S (sigmoidal)型(图 12-13)。

这种作图的一个有用的指数是 $C_o t$ 1/2，它表示 DNA 的半数再结合($f=1/2$)时的 $C_o t$ 值。大肠杆菌 DNA 的 $C_o t$ 1/2 约为 15 mol·s/L。而 T4 噬菌体 DNA 的约为 0.3 mol·s/L，这些数值说明大肠杆菌 DNA 的再结合(复性)比 T4 噬菌体 DNA 慢很多。大肠杆菌 DNA 的再结合为什么比 T4 噬菌体的慢？这是因为大肠杆菌的 DNA 比 T4 噬菌体的大，而且已知它们的

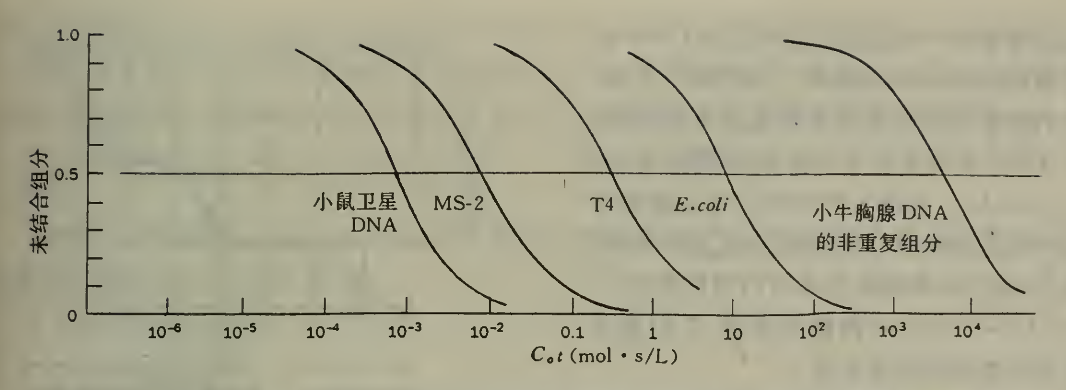


图 12-13 这些 C_0t 曲线描述了几种热变性 DNA 的再结合动力学

引自 R. J. Britten and D. E. Kohne, Science. 1968. 161 : 530.

DNA 都是单一序列(即没有或很少有重复序列),因而在用等量(注意:在实验中所用 DNA 的量是每升溶液中含有相同摩尔数的核苷酸)DNA 切成片段后,原来越长的 DNA,其不同类型的片段在溶液中的浓度就越低。由于大肠杆菌的 DNA 比 T4 噬菌体的 DNA 大,所以它的不同类型的 DNA 片段的浓度比较低,因而复性的速度就比较慢了。我们把每个单一序列所含的碱基对数称为该 DNA 的复杂度(complexity)。例如,一个重复序列(repeating sequence) $(AGCT)_n$,它的复杂度为 4,而一个大肠杆菌染色体是由非重复的 4×10^6 bp 序列组成的,它的复杂度即为 4×10^6 bp。由此可见,一个 DNA 的 C_0t 1/2 值越大,表明它的复杂度越高。

在用相同方法研究小鼠 DNA 时得到了出人意料的结果。由于哺乳动物的基因组比大肠杆菌的大 3 个数量级左右,因而其 C_0t 1/2 值似乎应该为 10^4 mol · s/L。一个 10^{-4}mol/L 的具有这种 C_0t 1/2 值的 DNA 溶液,将需 10^8 s(约 3 年)才能有一半再结合起来。然而令人惊异地是,小鼠 DNA 中有一部分(约占 10%)仅在几秒钟内就有一半再结合了。事实是,小鼠 DNA 的这一部分的再结合速度甚至比最小的病毒 DNA 还要快。这一事实说明,在小鼠 DNA 中有比较小的片段,它们在整个 DNA 中有很多拷贝(即重复序列),在切成片段时,它们的浓度很高,因而复性很快。

三、重复序列

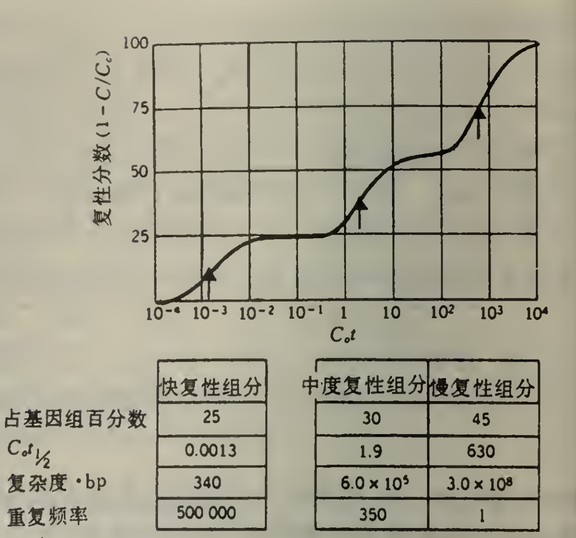
复性动力学的分析表明,病毒和原核 DNA 中没有或很少有重复序列,因为它们的 C_0t 曲线(如图 12-13 中 MS2,T4 和大肠杆菌的)是个简单的 S 型。与之相反,真核 DNA 的 C_0t 曲线表现为多曲折型,即有多个 C_0t1/2 值。由于这种曲线很复杂,而且不同真核生物的曲线也不相同,为了说明问题,图 12-14 所列是个模式图。通过对这种图的分析,可以估测出不同重复序列的重复频率,每种重复序列片段的长度范围,以及各种重复序列和单一序列各在总 DNA 中所占的百分数(见图中下面的表)。

根据对真核生物的 C_0t 曲线的分析可把它们的 DNA 相对地分为 4 类:①单一序列(unique sequence),每个单倍体基因组中含 1 拷贝(也包括含几拷贝的);②中度重复序列(moderately repetitive sequence),每个单倍体基因组中含$<10^6$ 拷贝;③高度重复序列(highly repetitive sequence),每个单倍体基因组含$>10^6$ 拷贝;④反向重复序列。

(一)反向重复序列

这是真核 DNA 中再结合最快的部分,约占某些基因组的 10%。复性动力学的研究表明,

它的复性为一级反应，说明这种 DNA 含有靠得很近的反向重复（自我互补）序列，它们能够自我折叠起来形成发夹结构（图 12-15）。反向重复序列的长度范围为 100～1 000 bp。这样大的序列不可能是在进化中随机出现的。用这种反向重复序列（分离出来的）对中期染色体进行原位杂交（in situ hybridization）的研究表明，它们是分布在染色体的许多点处。



	快复性组分	中度复性组分	慢复性组分
占基因组百分数	25	30	45
$C_0t_{1/2}$	0.0013	1.9	630
复杂度·bp	340	6.0×10^5	3.0×10^8
重复频率	500 000	350	1

图 12-14　真核生物 DNA 复性动力学模式图

在人类基因组中这种反向重复序列大约有 2×10^6 个，它们的功能尚不清楚。然而，由于用配对的折回结构形成的十字形结构（cruciform structure）的稳定性（图 12-15 b）仅比正常的双螺旋 DNA 稍微差一些，因而有人认为，染色质中的反向重复序列的功能可能是作为某种分子开关（molecular switch）。

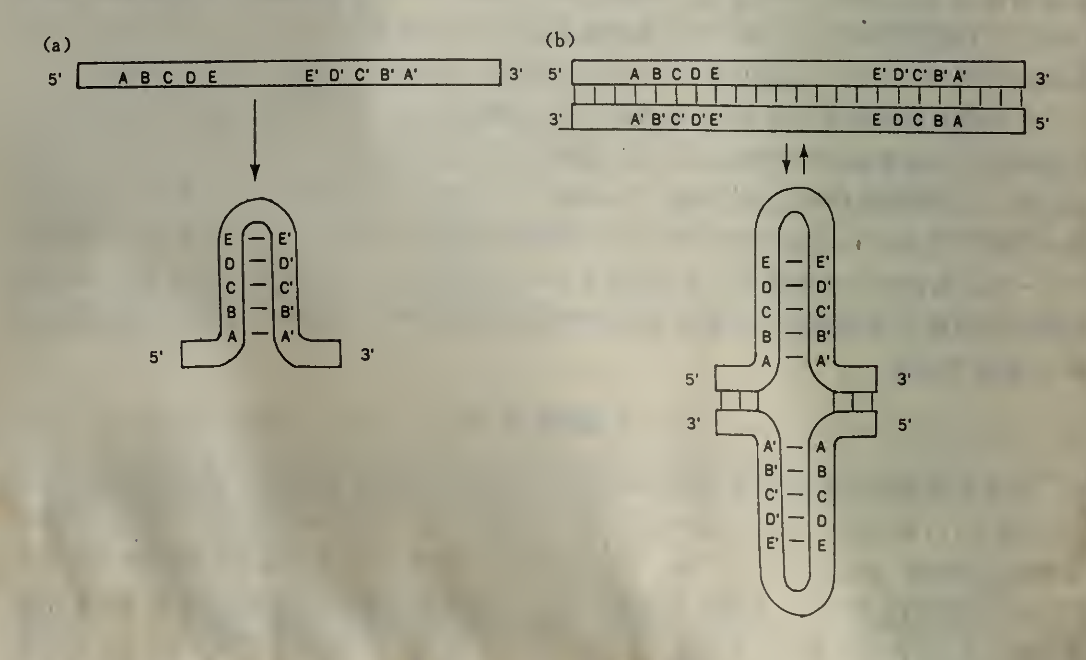


图 12-15　DNA 中的折回结构

(a)含有反向重复序列的单链 DNA，在复性条件下，形成一个碱基配对的环，叫做折回结构（fold-back structure）。这里 A 与 A′互补，B 与 B′互补，如此等等；(b)在双链 DNA 中的一个反向重复序列可设想它形成一种含有两个折回结构的十字形构象。这种结构的稳定性比双螺旋要差些，因为它有未配对的环。

（二）高度重复序列

此种序列是由长度不到 10 bp 的 DNA 片段所组成，在每个细胞的 DNA 中重复次数以百万计，在大鼠中这种序列约占其总 DNA 量的 10%。许多高度重复序列可用密度梯度离心法分

离出来（有的则不能），这是因为这些序列中富含 A，T，因而它们的浮力密度（buoyant density）比 DNA 总体较低。例如，将果蝇的 DNA 打碎后，在 CsCl 溶液中进行梯度离心后，其沉降图谱如图 12-16 所示。从图中可以看出，除了主带（1.700）外还有 3 条额外的带，称之为卫星（satellite）带（因而将它们的 DNA 称为卫星 DNA）。分析这些卫星带的 DNA 表明它们都是高度重复序列。

事实上，果蝇的这 3 种卫星 DNA 是分别由 3 种不同的（虽然是很类似的）7 核苷酸序列串联重复形成的，这 3 种序列是：

5′-ACAAACT-3′
⋮⋮⋮⋮⋮⋮⋮
3′-TGTTTGA-5′
卫星 Ⅰ

5′-ATAAACT-3′
⋮⋮⋮⋮⋮⋮⋮
3′-TATTTGA-5′
卫星 Ⅱ

5′-ACAAATT-3′
⋮⋮⋮⋮⋮⋮⋮
3′-TGTTTAA-5′
卫星 Ⅲ

卫星 DNA 也叫做简单序列 DNA（simple sequence DNA），据认为它们是不编码蛋白质或 RNA 的，许多高度重复的 DNA 是位于真核染色体的两个重要结构——着丝点（centromere）和端粒（telomere）处，这是用原位杂交技术证明的。

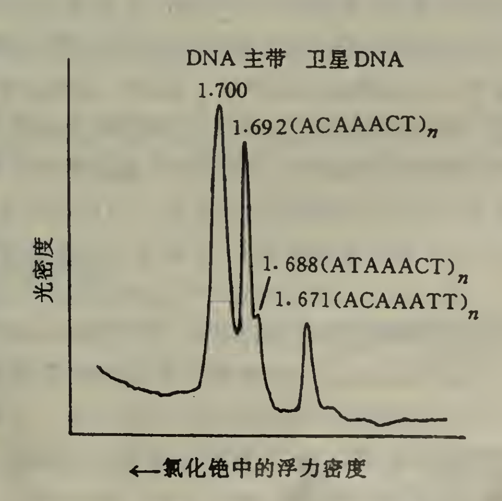


图 12-16　果蝇 DNA 在中性 CsCl 中离心至平衡时的沉降图谱

引自 Gall，J. G.，Cohen，E. H. and Atherton，D. D.

每个真核染色体都有一个着丝点，它的功能是作为那些（把染色体联结到有丝分裂的纺锤体上的）蛋白质的接合点。在细胞分裂时这种结合对于将染色体分配给子细胞是必需的。酵母染色体的着丝点已被分离并研究过（图 12-17）。着丝点功能所必需的序列约为 130 bp 长，它非常富含 A═T 对。高等真核生物的着丝点要比酵母的长很多。在高等真核生物中（不是在酵母中），卫星 DNA 普遍存在于着丝点区，并且是由数以千计拷贝的一种或几种短序列串连（一个挨一个，并且是相同方向）而成。已鉴定的卫星序列一般是 5～10 bp 长。卫星 DNA 在着丝点功能中所起的具体作用尚不清楚。

端粒是位于线型的真核染色体末端的序列，它有助于使染色体稳定。研究得最清楚的是简单真核生物的端粒。酵母端粒的末端有约 100 bp 的不太严格的重复序列，其形式为：

$$5'(\mathrm{T}x\mathrm{G}y)_n$$
$$3'(\mathrm{A}x\mathrm{C}y)_n$$

式中 x 和 y 通常在 1～4 范围内。产生这种重复序列的原因是，由于真核 DNA 为线型，而 DNA 聚合酶只能由 5′向 3′方向复制，因而在复制完毕时，在去掉其头上的 RNA 引物后，此一

段是无法补齐的。如果没有补救机制，则会使DNA每复制一次便会缩短一段，这当然是不可以的。补救的机制是由端粒酶(telomerase)完成的。此酶是个含有RNA的反转录酶，它以其自身的RNA为模板，连续不断地向DNA末端加上6个碱基的DNA序列，从而形成了端粒末端的重复序列。这样就保证了在复制过程中DNA不会越来越短了(详见第五章)。

虽然如上所述在构成果蝇的卫星序列的每个重复片段之间同源性很高，但并非所有生物的卫星序列都是如此。例如，已分析的一段小鼠的卫星序列中，其重复片段之间在序列上是有差异的。由小鼠中制得的此234 bp的片段(一个很长的卫星序列中的一小段)是由两个117 bp的序列串联而成的，但这两个序列之间有19%的差异(同源性81%)。进一步分析这两个序列发现，它们每一个都是由两个更小的片段串联重复组成，这4个片段之间的差异更大一些。再进一步分析表明，整个序列实际上是由一个9 bp的单位串联重复而成的。据推测，产生这种结果的原因可能是，最初的一个9 bp片段经过扩增而成为多重复单位，这些单位原本是一样的，但经过突变(包括碱基替换、插入和缺失)，它们彼此间就产生了差异。此重复序列再扩增而构成由更大单位形成的重复序列，这些序列又发生突变，如此反复而形成了现在我们看到的卫星序列。卫星序列之所以易于发生突变，是由于它们不编码蛋白质之故。而且这种重复序列在复制过程中，由于基因重组的结果，容易发生重复次数的改变。

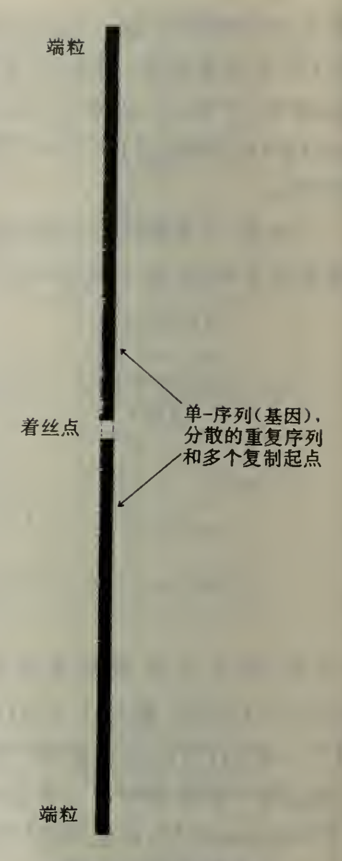


图12-17　一个酵母染色体的重要结构特征

在人类(和其它动物中)普遍存在着与卫星序列类似，由短的单位串联重复组成的序列，但它们的总长度短得多，大致由5～50个重复单位组成，称之为小卫星(minisatellite)或数目可变的串联重复(variable number tandem repeat，VNTR)序列。由于它们的重复次数变动很大，因而在一个群体中，在基因组的相同位点，不同个体的小卫星序列的长度变动很大，表现出高度的多态性。而且已经发现，在人类和动物中存在有小卫星家族，它们有共同的核心序列(core sequence)。因此，当我们用限制性内切酶将某一个体的DNA切断，经电泳将不同长度的DNA片段分开后，用制得的小卫星探针与之杂交，即可显示出许多条长短不等的带，这些带代表着该个体基因组中许多位点处的小卫星片段长度。这种图型被称为DNA指纹图(DNA finger printing)。之所以叫这个名字是因为它和人的指纹类似，能够用以区分每个个体。这是易于理解的，因为同一位点的每个个体的小卫星长度已经是高度变异的，而两个个体的多个位点(通常是显示出20～30个位点)小卫星序列的长度完全一样的可能性几乎是没有的。因此，DNA指纹图已在法医上得到了广泛地应用。

(三)中度重复序列

中度重复序列是由100到几千个bp的片段重复而成，它们分散在整个高等生物的基因组中，居于比它们更长的单一序列包括单拷贝和仅几拷贝的片段，约占基因组DNA总量的40%～70%，大多数基因在这类序列中间。在这类重复序列中，有些是由一组基因串联重复所组成，这些基因规定的产物是细胞大量需要的，例如，核糖体RNAs，tRNAs和组蛋白等。这些基因

的结构将在下面讨论。在这种DNA中鉴定得最清楚的是Alu家族(Alu family),这样命名是因为这种约为300 bp长的片段中,大多数片段含有一个限制性内切酶Alu I的酶切位点。Alu家族是人类基因组中存在最广泛的中度重复DNA,含有30万~50万个Alu序列,约为人类基因组DNA总量的1%~3%。它们显然是由一个编码7 SL RNA的基因衍生出来的。7 SL RNA是一种叫做信号识别颗粒的组分(见第八章),此信号识别颗粒在蛋白质合成中起作用。然而,Alu序列缺乏7 SL RNA基因中的一些部分,因而产生不出有功能的7 SL RNA。在猴和啮齿动物中也有Alu DNA,而在亲缘关系很远的生物,例如棘皮动物、两栖动物和禽类中则含有Alu样(Alu-like)序列。虽然在许多生物中Alu家族是最突出的中度重复DNA,但它并不是唯一的。序列分析已经证明,在脊椎动物的基因组中通常含有几种不同的中度重复DNAs。

除了表达产生核糖体RNAs,tRNAs和组蛋白的序列外,其余大多数不表达的中度重复DNA的功能是什么?从它们的片段长度和拷贝数看似乎是应该有不同的功能,然而,虽然已经提出了一些设想,但都没有得到实验的证实。在这些设想中觉得最有可能的是中度重复DNA的功能是作为调控序列,它们参与其附近基因的活化。另一种设想是中度重复DNA的某些家族为复制起点。第三种设想的中度重复DNA的功能是它们提高了真核基因组进化方式的多样性,这是通过促进染色体的重排(rearrangement),或者是它们作为库存,以便由其中得到新的功能性序列而实现的。诚然,有些中度重复序列在结构上有一定转座因子(transposable element,见第六章)的性质。然而由于所有这些设想都未能得到实验的证实,因此关于这些中度重复序列的功能可以说是迄今毫无所知。

由于对于这些中度重复序列的种种设想的功能都得不到证实,因而另一种设想,即它们可能就是废弃的无用的DNA,是在进化过程中残留下来的设想,也是需要认真对待的。当然是否是这样,也还需要用实验来证实。

四、串联重复的基因簇

在生物的单倍体基因组中,大多数基因仅出现一次(单拷贝),即使是那些需要量很大的蛋白质的基因也是单拷贝。然而rRNA,tRNA和组蛋白的基因是多拷贝的。

(一)rRNA基因组织成一套并重复

我们已经知道,在大肠杆菌的基因组中,尽管其它基因都是单一序列的,但rRNA和tRNA的基因却是多拷贝的。在真核生物中,规定18 S,5.8 S和28 S的基因总是按18 S,5.8 S,28 S的顺序排列,在这些基因之间有短的被转录的间隔区(spacer)将它们分开,形成一个约7 500 bp的转录单位(图12-18)。此基因簇(gene cluster)的初始转录物是个45 S RNA,由此RNA经转录后加工形成成熟的rRNA。这个转录的基因簇是串联重复的,在簇与簇之间用不转录的间隔区分开。这些串联重复序列典型的长约12 000 bp,尽管不同品种生物的不转录间隔区的长度有所不同,而且同一生物的基因与基因之间的不转录间隔区的长度也有比较小的差异。用放射性标记rRNAs进行杂交的

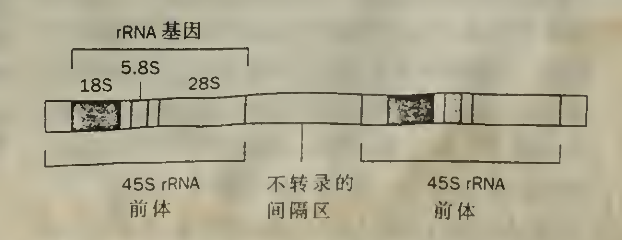


图12-18 编码18 S,5.8 S和28 S rRNA的基因组成一个转录单位,此转录单位串联重复,在两个转录单位之间有个不转录的间隔区

定量测定表明，这些 rRNA 基因(rDNA)是分布在几个染色体中，其在单倍体中的数目因品种的不同而不同，在少于 50～10 000 个以上的范围内变动。人的 rRNA 基因簇为 50～200 个，分布在 5 个以上染色体上。

在典型的间期细胞核中 rDNA 浓缩形成一个核仁。在核仁中这些基因迅速不断地由 RNA 聚合酶 I 转录。放射性标记实验表明，核仁也是这些 rRNAs 转录后加工的场所，并且在胞浆中合成的核糖体蛋白质也到核仁处与这些 rRNA 装配形成未成熟的核糖体亚基。仅当这些未成熟的核糖体亚基被运输到胞浆中后，才最后装配成成熟的核糖体亚基，这大概是为了避免在细胞核中翻译那些部分加工了的 mRNA 之故。

和上述几种 rRNA 的基因相似，编码 5 S rRNA 的基因也是安排成簇的，它总共含有几百到几十万个串联重复，分布在一个或更多个染色体中。当前研究得最清楚的是爪蟾(*Xenopus laevis*)的 5 S rRNA 基因，它的重复单位是由一个 5 S rRNA 基因，一个邻近的假基因(pseudogene，5 S rRNA 基因的一个 101 bp 的片段，它是不被转录的)和一个不转录的间隔区所组成(图 12-19)。间隔区的长度不相同，但平均约为 400 bp。5 S rRNA 基因是在核仁之外由 RNA 聚合酶 Ⅲ 转录的，所以 5 S rRNA 必须被转运到核仁中以便掺和到核糖体的大亚基中。tRNA 基因也是多重复的和成簇的，但对它们的结构所知甚少。tRNA基因是由 RNA 聚合酶 Ⅲ 转录的。

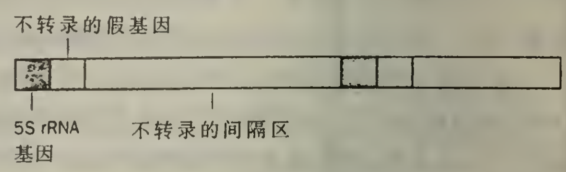


图 12-19　爪蟾 5 S rRNA 基因的结构

(二)组蛋白基因是成簇的并串联重复许多次

编码蛋白质的基因是怎样安排的?组蛋白基因是最早被分离并鉴定的，这是因为在迅速分裂的海胆胚细胞中组蛋白 mRNA 的含量非常丰富之故。这些海生的无脊椎动物由一个受精卵发育成 1 个1 000个细胞的囊胚(blastula)是在约 10 h 完成的，在此期间必须合成大量的组蛋白，以便形成新的染色质。诚然，在早期胚胎发生时期所合成的蛋白质中有 1/4 以上是组蛋白。而且在此期间所合成的 mRNA 中约 70%是组蛋白 mRNA，因而它们是易于分离出来的。用提取的组蛋白 mRNAs 与海胆的 DNA 杂交并分析其动力学，从而测得其组蛋白基因的拷贝数。结果发现，其杂交的速度比单拷贝序列的速度快几百倍，说明组蛋白基因是高度重复出现的。在海胆的几个品种中，组蛋白基因的拷贝数在 300～1 000 范围内(表 12-4)。在其它生物中的拷贝数低一些。例如，酵母中每种组蛋白基因仅为 2 拷贝。一种生物的组蛋白基因拷贝数与其对组蛋白 mRNA 合成速度的需求是相关的。

表 12-4　组蛋白基因的重复频率

品种	基因的拷贝数
海胆	300～1 000
果蝇(*Drosophila melanogaster*)	110
爪蟾(*Xenopus laevis*)	20～50
小鼠	10～20
鸡	10
人	30～40

海胆的 5 种主要组蛋白的基因是在一个 7 kb 的重复单位中成簇存在的(图 12-20)。在此重复单位中 5 个编码区与 5 个不转录的间隔区交替出现。这一 5 种组蛋白基因簇串联重复很多次。在不同组织中以及发育的不同时期表达不同套的组蛋白基因。在果蝇中(图 12-20)这 5 种组蛋白的基因也是成簇在一起并串联重复。据认为，这种安排可通过一套组蛋白基因的同步转录而灵敏调控组蛋白的合成。然而在某些脊椎动物中，几种组蛋白的基因是分散开的，说明

为了协调基因的表达，这些基因并不一定必须成簇。

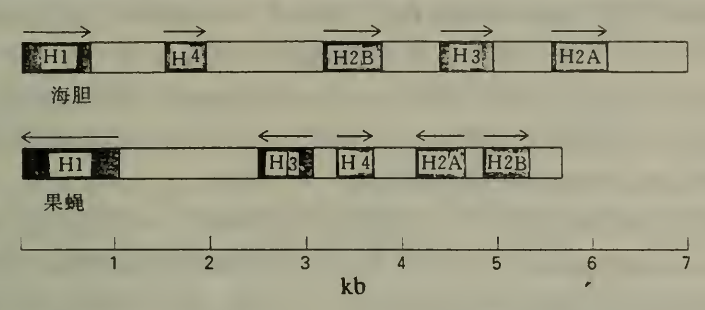


图 12-20 在一种海胆(*Strongylocentrotus purpuratus*)和一种果蝇(*Drosophila melanogaster*)中组蛋白基因簇的图谱

箭头表示转录的方向。

组蛋白基因的表达还有两个值得注意的特点：①组蛋白基因没有内含子。在脊椎动物中这种连续的基因是罕见的；②组蛋白 mRNA 没有多聚 A 尾巴。没有内含子和多聚 A 尾巴大概是能使组蛋白 mRNA 可以很快地生成，并且很快地转运到胞液中。

五、许多主要蛋白质是由单拷贝基因编码的

前面已经谈到核糖体 RNA 和组蛋白的基因是重复多次的。那么编码大量产物的其它基因是否也是多拷贝的？Donald Brown 在重组 DNA 技术出现之前就解答了此问题。他由家蚕(*Bombyx mori*)中分离出丝的丝心蛋白(fibroin) mRNA。选择这一体系进行研究是因为丝心蛋白 mRNA 分子具有结构特点而易于分离，而且在幼虫发育的特定阶段其单一类型的巨细胞(giant cell)能合成非常大量的丝心蛋白。用提纯的丝心蛋白 mRNA 与基因组 DNA 杂交的结果发现，在每个单倍体基因组中仅有一个丝心蛋白的基因。

此结果对真核生物基因的表达具有重要意义。它表明即使仅有一拷贝的基因即可合成出大量的由它编码的特定蛋白质。丝心蛋白的一个基因是合成 10^4 个 mRNA 分子的模板，这些 mRNA 分子可稳定地存在几天。每个 mRNA 又是合成 10^5 个蛋白质的模板。可见一个基因即足以合成出 10^9 个蛋白质分子。

核糖体 RNA 和组蛋白基因的多拷贝是特殊的，不是普遍的现象。丝心蛋白的基因只有一个是许许多多编码蛋白质的基因的一个例子，即使是产量很大的蛋白质的基因也是一个。例如，网织红细胞中含有一个或很少拷贝的血红蛋白亚基的基因。同样，卵清蛋白(ovalbumin)是卵清中的一种主要蛋白质，它的含量很大，是由产蛋母鸡的输卵管合成的，而在输卵管中每个单倍体基因组中仅有一拷贝卵清蛋白基因。

六、基 因 扩 增

特定的一套基因的选择性复制过程称为基因扩增(gene amplification)，在正常情况下，它仅发生在某些生物的生命过程中的特异阶段。

(一)正在发育的卵母细胞中 rRNA 基因扩增了

在胚胎发生的早期阶段，蛋白质合成的速度非常快，致使某些生物的正常基因组中所含 rRNA 基因的数量不能满足需要。在这些生物中最明显的有某些昆虫、鱼类和两栖动物，在它

们的正在发育的卵母细胞(未成熟的卵细胞)中,rRNA 基因被选择性地复制了。此扩增过程的一个最令人惊异的例子是爪蟾的卵母细胞,在其中 rRNA 的基因扩增至约为爪蟾体细胞的 1 500倍,即产生了约 2×10^6 套 rRNA 基因,使细胞总 DNA 的近 75%为 rRNA 基因。扩增的 rRNA 基因以凸出在染色体之外的环状物(extrachromosomal circle)形式存在,每个环状物含有一或两个转录单位,它们被组织在数以百计的核仁之中。因此,在成熟的爪蟾卵母细胞中含有约 10^{12}个核糖体,此数为大多数爪蟾细胞中的 200 000 倍。

rRNA 基因扩增的机制是怎样的?一个重要的线索是发现,一个给定的凸出在染色体外的核仁(extra-chromosomal nucleolus)中的所有不转录间隔区的长度都是一样的,然而我们已经谈到,其相应染色体中的间隔区则是长短明显不一致的。此发现提示,在单个核仁中的所有 rRNA 基因环都是由一个染色体基因传下来的。已证明基因的扩增分两个阶段:第一阶段是低水平的扩增,随后是第二阶段的大量扩增。因此,很可能是在第一阶段仅有很少几个染色体的 rRNA 基因被一个尚不清楚的机制复制了,并且新生成的子链 DNA 被放出来形成在染色体之外的环。然后在第二阶段这些环被大量地复制。

(二)卵壳(chorion)基因的扩增

唯一已知的其它按程序扩增的基因是果蝇卵巢滤泡(follicle)细胞中编码卵壳蛋白类的基因。在卵壳生成之前每个卵巢滤泡细胞的整个单倍体基因组复制 16 倍,然后只有卵壳基因被选择性复制约 10 倍。有意思的是在丝蛾(silk moth)卵母细胞中的卵壳基因不发生扩增,而是在它的基因组中有多拷贝的卵壳基因。

(三)在选择压力(selective pressure)下基因的高度扩增

在用化学疗法治疗癌症时观察到一个普遍的现象,即在连续投给对细胞有杀伤作用的药物时,开始时对肿瘤很有作用,然后引起肿瘤对药物的抗性逐渐增加,从而达到使药物失去治疗作用的地步。用细胞系(cell line)所做的研究表明,一个细胞系能获得这种药物抗性的机制之一是它过量地产生此药物的靶酶(target enzyme)。能够观察到这种过程的一个例子是用氨甲蝶呤(methotrexate)处理培养的动物细胞。氨甲蝶呤是二氢叶酸的类似物,因而它是二氢叶酸还原酶(DHFR)的竞争性抑制物。由于 DHFR 是脱氧胸苷酸合成的一个关键酶,而氨甲蝶呤能够强烈地抑制此酶的活性,从而封阻了脱氧胸苷酸的合成,所以氨甲喋呤能杀死快速复制的细胞(包括癌细胞),是一种治疗许多癌症的有效药物。当用氨甲蝶呤处理培养的动物细胞,而氨甲喋呤的剂量缓慢增高时,可得到存活下来的细胞。在最后存活下来的细胞中含有高达 1 000拷贝的 *DHFR* 基因,从而能够产生大大过量的二氢叶酸还原酶。

这种基因扩增的机制尚不清楚,但此发现说明了基因组的动态本质。基因的扩增是经常发生的,而选择压力(用氨甲喋呤处理)能使这种扩增保留下来。由于对二氢叶酸还原酶的这一工作揭示了长期用低于致死量的剂量处理细胞时可使细胞产生抗药性,因而对用化学疗法治疗癌症的策略也是有用的。

这种基因的扩增是遗传不稳定的。当细胞在没有氨甲蝶呤的条件下进一步生长时,额外的 *DHFR* 基因便会逐渐消失掉。

七、割裂基因(split gene)

在真核生物中,大多数编码蛋白质的基因是不连续的,即在其编码氨基酸的序列之间插入了不编码的序列,故称之为割裂基因。

(一)割裂基因的发现

在细菌中,多肽链是由DNA中一排连续的密码子编码的。在许多年里人们认为高等生物的基因也是连续的。这个概念在1977年被出乎意料地打破了,因为有几个实验室的研究者发现许多基因是不连续的。例如,血红蛋白β链基因中编码氨基酸的序列之间被一个550 bp长的非编码序列和一个120 bp短的非编码序列分隔开了(图12-21)。这样,β-珠蛋白基因被割裂成为3个编码序列。这种出乎意料的结构是用电镜观察mRNA与含有编码它的基因的DNA片段的杂交物时发现的。图12-22是这种杂交的模式图。图中表明DNA的双链有些部位拆开了,使得其中的一条与mRNA互补的链和mRNA杂交了;而另一条DNA单链则凸出成环。如果基因是连续的,则在电镜下只能看到一个单链(较细)的环(图12-22 A)。如果基因中含有一个间插序列(图12-22 B),则可看到两个单链的环和一个较粗的双链DNA的环(由间插序列所形成)。用小鼠β-珠蛋白mRNA与含有β-珠蛋白基因的DNA片段进行杂交,在电镜下观察(图从略)的结果,清楚地表明了在该基因中有两个间插序列。

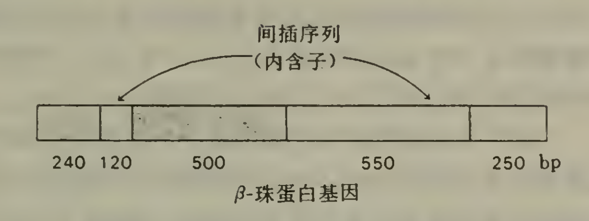


图12-21 β-珠蛋白基因

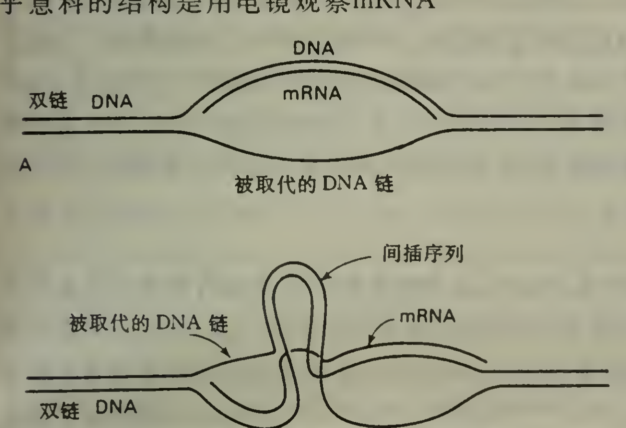


图12-22 用电镜检测间插序列的示意图

已证明,在转录时这些不编码的间插序列是和编码的序列一起转录的,从而产生一个比成熟mRNA大得多的mRNA前体(或称初始转录物),然后经过一个剪接(splicing)过程,将不编码的序列去掉,把编码的序列连在一起,成为成熟的mRNA,它才是合成蛋白质的模板。因此,把割裂基因中编码的序列称为外显子(exon,因为它们是被表达的区段),而把非翻译的间插序列叫做内含子(intron)。或者说内含子是在形成成熟mRNA分子时被剪接掉的序列。

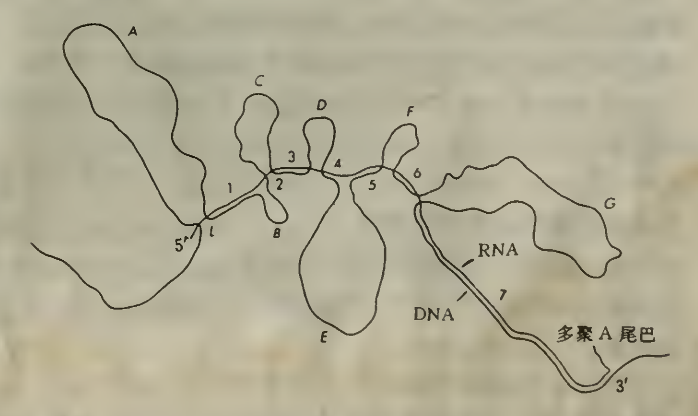


图12-23 一个割裂基因的结构

将含有卵清蛋白基因的DNA与卵清蛋白mRNA进行杂交。基因的8个外显子(L和1～7)与mRNA中的互补区段杂交在一起,7个内含子(A～G)则打成环。

现在看来，在几乎所有脊椎动物的基因中都有内含子。在真核微生物(如酵母)的基因中也有内含子，不过其存在的频率要比高等生物低得多。不同基因所含内含子的数目以及内含子的大小是很不相同的。编码胶原蛋白(collagen)长多肽链的基因含有近 40 个内含子。而且一个基因的内含子的核苷酸数常常比编码的外显子的核苷酸数多很多(图 12-23)，这个事实至少部分地解释了为什么初始转录产生的 RNA 很大。初始产生的许多 RNA 很大的现象是早就知道的，但长期以来不明白它们为什么这样大。

并非所有的高等生物的基因都含有内含子。在已经分析过的基因中，已知(也仅仅是)编码组蛋白和干扰素(interferons)的基因是不含内含子的。不过我们可以用重组 DNA 技术制备出编码任何蛋白质的不含内含子的基因。例如，可以通过对某一 mRNA 进行反转录而获得与该 mRNA 互补的 DNA(complementary DNA，简称 cDNA)。通常把与某一蛋白质 mRNA 互补的 DNA 称为该蛋白质的 cDNA 基因，而把存在于基因组中的含有内含子的基因称为该蛋白质的基因组基因，以示区别。含有内含子的真核基因组基因只能在真核细胞中进行表达，不能在原核细胞(例如大肠杆菌)中表达，因为原核细胞没有将内含子剪接掉的能力。cDNA 基因则既可在真核细胞中表达(虽然由于尚不清楚的原因，其表达效率常常比基因组基因差一些)，也可在原核细胞中表达出正确的多肽链来。

(二)内含子的意义

已分析的 DNA 序列的数目的迅速增多，揭示出在原核生物的结构基因中内含子是罕见的，在低等真核生物如酵母中内含子不很普遍，而在高等真核生物中则大量存在(在脊椎动物的结构基因中仅知编码组蛋白和干扰素的基因没有内含子)。在大多数割裂基因中外显子的长度均在 100～250 bp 范围内。与之相反，内含子的长度则在 50～20 000 bp 以上范围内。而且在一个基因中内含子的数目可以多得惊人，迄今已知的内含子数目最多的是编码鸡 α2(1)胶原蛋白的基因，在此 38 kb 的基因中有 51 个内含子。因此，在一个典型的脊椎动物的结构基因中不表达的序列约占 80%。

内含子的功能是什么?认为所有内含子都是无用的 DNA 的观点看来是站不住脚的。因为如果只是无用的 DNA 序列，则很难解释为什么由于内含子的存在还发展出来了剪接的机构(见第四节)，而在进化中发展出剪接的机构并不比简单除掉内含子有任何明显的优越性。然而，在大多数基因中内含子又确实没有什么明显的功能(尽管内含子可以保护基因家族中基因的完整性，见后；以及在某些基因的转录中作为调控成分)。而且事实上，编码某一蛋白质的基因中所含内含子的数目在不同生物中可以是不同的，甚至在同一个体中也可以不同。例如，大鼠有两个功能的胰岛素基因，其中之一有两个内含子，这和大多数啮齿类动物的胰岛素基因是一样的，而在另一个基因中则只有一个内含子，另一个内含子缺失掉了。因此有人认为，在遗传进化的一个早期阶段内含子具有重要的功能，后来就不再重要了，就是说内含子可能是遗传的残留物。

Walter Gilbert 提出了一个关于内含子作用的假说。他认为编码蛋白质的基因是由外显子的集成产生的，这种集成过程是通过内含子序列之间的重组而将外显子聚集在一起的。因此，现在看到的内含子是促使蛋白质进化过程的残迹。有许多实验证据支持此假说，举例如下：

(1)鸡 α2(1)胶原蛋白的三股螺旋(triple helix)区是由 332 个重复的三联体 Gly-x-y(Gly 代表甘氨酸，x 和 y 代表其它不固定的氨基酸)组成的，它由胶原蛋白基因 52 个外显子中的 42 个外显子所编码。这 42 个外显子的碱基对数都是 9 的倍数(其中有 23 个外显子的长度为

54 bp，其余的为 45，99，108 或 162 bp 长），每个外显子都以甘氨酸的密码子开始，以 y 的密码子告终。这种分布提示，编码胶原蛋白三股螺旋区的基因片段是由用内含子夹起来的基因元件(genetic element)经多次加倍进化而成的。

(2)根据鸡肌肉中丙酮酸激酶基因的碱基序列分析此酶的结构发现，此基因共有 10 个外显子，每个外显子都编码蛋白质二级结构的一个独立成分，大多数内含子都标志出多肽链形成一个回折(reverse turn)的位置。可见，丙酮酸激酶的基因显然是通过一系列较小的编码蛋白质的单位连接在一起形成的，并且利用 RNA 的剪接把它们表达成为一条多肽链。

(3)对于 Gilbert 的假说最有说服力的证据是对低密度脂蛋白(low density lipoprotein，LDL)受体基因序列的分析。LDL 受体由 839 个氨基酸残基组成，是整合在质膜上的一种蛋白质，它的功能是结合 LDL，并通过胞吞作用(endocytosis)将 LDL 转运到细胞内。LDL 受体的基因长 45 kb，含有 18 个外显子。大多数外显子都是分别编码蛋白质的一个功能结构域(functional domain)。而最令人感兴趣的事实是，在它的外显子中有 13 个外显子编码的多肽片段与其它蛋白质中的片段是同源的(homologous)。例如，它的某些外显子编码的多肽片段在补体 C9(complement C9)中也存在等。很明显，LDL 受体的基因是由外显子按标准构成的，而这些外显子也编码其它蛋白质的一部分，许多其它真核蛋白质的组成也与此类似。

然而由此提出的一个问题是，如果 Gilbert 的假说是对的，即内含子是原初基因重新组合过程的残迹的话，那么为什么在低等生物中没有或几乎没有内含子呢?而高等生物是由这些低等生物进化而来的呀? 例如，酵母丙酮酸激酶的氨基酸顺序与鸡的同源性为 45%，但酵母丙酮酸激酶的基因中却没有内含子。看来对此问题的最为合理的解释似乎是：低等生物的生命类型是助长效率的，它的内含子已选择性地缺失掉了。与之相反，高等生物已经适应了稳定的环境，使它们缺失掉内含子的选择压力已经大大减少了(尽管如前所述，大鼠胰岛素基因提供了一个明显的将内含子缺失掉的例子)，因而存留了那样多的内含子。

内含子除了使新蛋白质的进化易于发生外，在基因家族中的内含子可能还有保护作用，即保护基因家族中相邻近的外显子不被不等交换(unequal crossing-over)所消除。双份的基因特别易于进行这种形式的降解，因为它们的碱基序列相似促进了它们的错配(mispairing)。然而，由于这些外显子与可变性大得多的并因而不易错配的内含子交替存在的结果，抑制了此过程。原核生物和酵母的基因家族很少，因而它们不太需要这种保护作用。

八、成簇的基因家族：血红蛋白基因的结构

在一个生物个体内，只有少数蛋白质确实是独特的。然而许多蛋白质，例如，消化酶类(胰蛋白酶、胰凝乳蛋白酶和酯酶)或者各种不同的胶原蛋白等，它们分别构成在结构和功能上相关联的蛋白质家族，每种蛋白质都是某一家族的成员。我们把编码一个蛋白质家族(也包括在结构和功能上相关联的 rRNA 和 tRNA 等)的若干个基因称为一个基因家族(gene family)。一个基因家族的成员彼此在结构和功能上是相似的，但也包括在结构上相似而没有功能的基因，没有功能的基因称为假基因。一个基因家族的基因常常是成簇地存在于一个染色体上，但也有分散的，甚至位于不同染色体上。

根据基因家族的情况目前可分为三种类型，即：①简单多基因家族；②复杂多基因家族；③由发育阶段控制的多基因家族(图 12-24)。5 S rRNA 基因家族属于简单的多基因家族，其中的基因结构是相同的，基因与基因之间由中度重复序列隔开。复杂多基因家族的各个成员并

不都是相同的。例如,海胆和果蝇的 5 个组蛋白基因构成为一个复杂的多基因家族,这 5 个组蛋白的基因作为一个单位可重复多达近千次,每个基因单独地按一定方向转录。又如果蝇的 tRNA 基因是由多种不同氨基酸的 tRNA 基因集中在一起而构成复杂的多基因家族的。

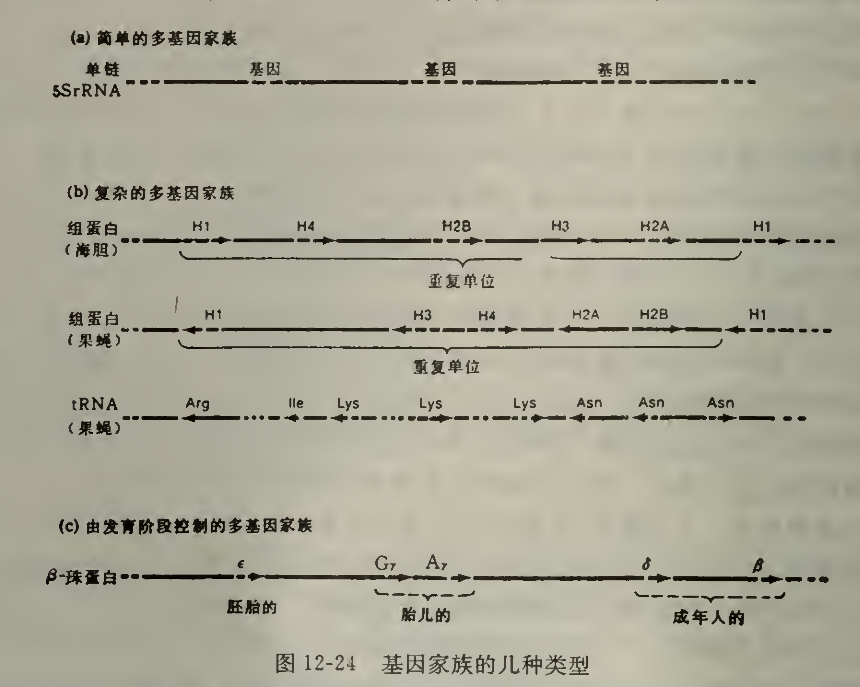


图 12-24　基因家族的几种类型

人类珠蛋白基因家族是典型的由发育阶段控制的多基因家族。关于 rRNA,tRNA 和组蛋白基因的结构前面已经讨论了。下面介绍人类珠蛋白基因的结构。

(一)人类血红蛋白的基因是按发育的次序安排在两个簇中

血红蛋白的研究使我们获得了许多关于蛋白质的结构和功能的知识。同样,研究血红蛋白基因及其表达方式,曾经是洞察一般真核基因功能的来源。例如,研究 β-珠蛋白 mRNA 与 β-珠蛋白基因的杂交导致发现了真核基因是用内含子断开的。而且,编码人类血红蛋白两类亚基的基因家族是迄今鉴定得最清楚的两个成簇的基因家族。

表 12-5　人类血红蛋白的亚基

胚胎的血红蛋白	
Hb Gower 1	$\zeta_2\varepsilon_2$
Hb Gower 2	$\alpha_2\varepsilon_2$
Hb Portland	$\zeta_2\gamma_2$
胎儿的血红蛋白	
HbF	$\alpha_2\gamma_2$
成年的血红蛋白	
HbA	$\alpha_2\beta_2$
HbA_2	$\alpha_2\delta_2$

成年人的血红蛋白主要是血红蛋白 A (HbA),约占总血红蛋白的 97%,它是由两条 α 链和两条 β 链组成的四聚体($\alpha_2\beta_2$)。占总血红蛋白 2%的是血红蛋白 A_2(HbA_2),它由 2 条 α 链和 2 条 δ 链所组成,所以 HbA_2 的组成为 $\alpha_2\delta_2$(表 12-5)。另外还有约 1%的 HbF。然而,在人类胚胎中最初生成的血红蛋白是个 $\zeta_2\varepsilon_2$ 四聚体(称为 Hb Gower 1),其中 ζ 和 ε 分别为 α-样和 β-样亚基。在怀孕后 8 周左右,胚胎的血红蛋白亚基(在新生成的红细胞中)被 α 亚基和 β-样的 γ 亚基所取代,从而生成胎儿的血红蛋白(HbF),$\alpha_2\gamma_2$(在此转变期间存在有 $\alpha_2\varepsilon_2$ 和 $\zeta_2\gamma_2$ 型血红蛋白,它们被分别命名为 Hb Gower 2 和 Hb Portland)。由出生几周前开始,γ 亚基逐渐为 β 亚基所取代,最后形成成年的血红蛋白。血红蛋白各种亚基在胚胎中的变化见图 12-25。

在哺乳动物中，编码血红蛋白的 α-样和 β-样亚基的基因分别形成两个不同的基因簇，并存在于不同的染色体上。这种分布主要是通过血红蛋白基因的克隆和进行序列分析确定的。这些基因是用Southern 印迹法(Southern blotting)由基因组文库中筛选出来的。在筛选过程中所用的探针是由血红蛋白 mRNA 制成的，由于在网织红细胞中的 mRNA 主要是血红蛋白 mRNA，因而易于将它分离出来。在人类以及许多其它哺乳动物中，这两个簇的基因都是按其在发育过程中的表达次序由 5'→3' 排列在编码链上(图 12-26)，即在 α 簇中，胚胎的 ζ 链基因在两个 α 链($\alpha2$ 和 $\alpha1$)基因的前面，$\alpha2$ 和 $\alpha1$ 分别是胎儿和成年血红蛋白的成分；在 β 簇中，胚胎的 ϵ 链的基因在最前面，然后是胎儿的两个 γ 链的基因，最后是成年的 δ 和 β 链基因。由此可见，人类珠蛋白基因的顺序是与它们在发育过程中的表达次序相吻合的。这种次序在哺乳动物中是普遍的，但也并非都是这样。例如，在小鼠的 β 基因簇中，成年的基因在胚胎基因的前面。

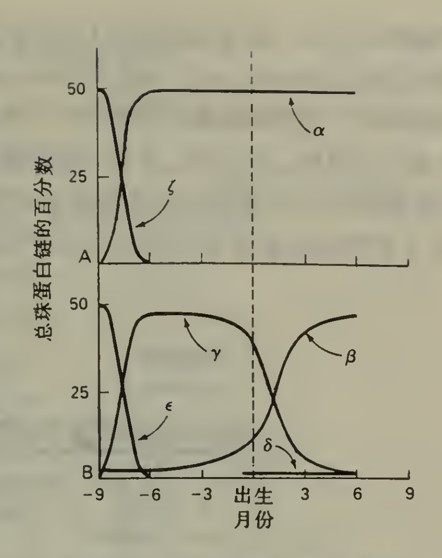


图 12-25　在人类发育过程中血红蛋白基因的表达

A. α 和 ζ 基因；B. β、γ 和 δ 基因。

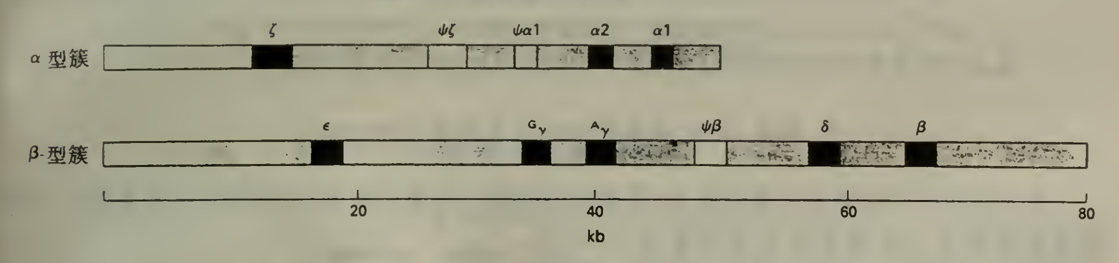


图 12-26　人类珠蛋白基因在它们各自的编码链上的安排

在 α 和 β 基因簇中都有假基因。这些序列与其毗邻的基因是同源的，但它们不编码功能的产物。前缀 ψ 指明是一个假基因，例如 $\psi\alpha1$，则表明它是与 $\alpha1$ 基因相类似的假基因。这种假基因的几种结构特点使得它们不活动：(1)缺乏剪接的 5'共有序列(见第四节)；(2)多聚 A 化信号由 AATAAA 突变为 AATGAA；(3)正常的起始密码子 ATG 为 GTG 所取代；(4)由第 38 号氨基酸的密码子开始缺失了 20 个核苷酸，从而产生了 3 个终止密码子，阻止了多肽链的完整合成。为什么会有假基因？它们中的大多数是基因的残遗物，是在基因进行重复时其序列有所漂移而产生的，并且变成了不活动的基因。由此我们可以看到，真核基因组既有其强有力的一面，也有被摧毁的一面。

(二)所有血红蛋白基因都有相同的外显子-内含子结构

无论在 α 珠蛋白基因簇或 β 珠蛋白基因簇中，编码蛋白质的序列都不到其全长的 5%。造成这种情况的原因，主要是由于在每个簇中都存在有不同的把基因分开的间隔区的结果。此外，所有已知的脊椎动物(哺乳类、爬虫类、禽类、两栖类和多骨鱼)的珠蛋白基因，包括肌红蛋白的基因和血红蛋白的假基因，都有 3 个位置相似的编码序列(外显子)，它们被两个稍有不同的不表达的间插序列(内含子，图 12-27)分隔开。这种基因结构显然是在脊椎动物发展史的很

早期即已形成。而这些基因的 DNA 序列之间的差异提示它们是由一个祖先的血红蛋白的基因经过重复(duplication)和趋异(divergence)而产生了 α 和 β 珠蛋白基因。由一种原始的脊椎动物的单一类型血红蛋白亚基变为 α 和 β 两种类型的亚基大致发生在 5 亿年前，即在脊椎动物进化的早期。由于肌红蛋白与血红蛋白的 α 和 β 亚基的差异比血红蛋白的 α 和 β 亚基之间的差别还大，所以肌红蛋白与血红蛋白的趋异应在 α 和 β 基因分开之前。在 α 和 β 基因分开之后，它们再经过重复和多样化，从而产生了当前人类所具有的 α 和 β 两个基因系列。

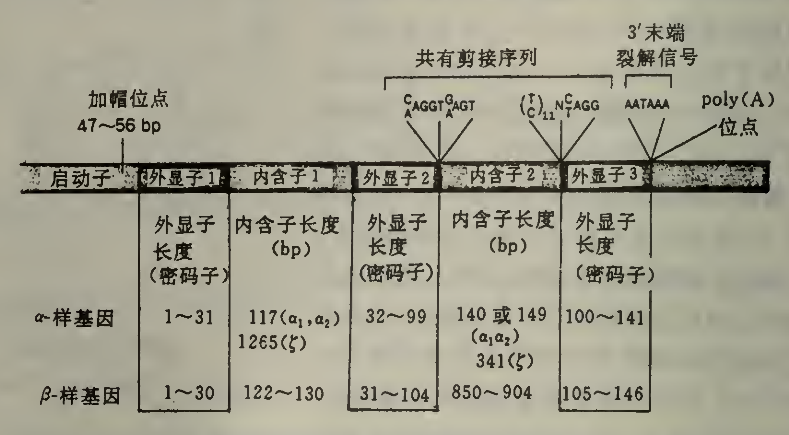


图 12-27　典型血红蛋白基因的结构

指明了外显子-内含子边界处的保守序列(剪接序列)和基因 3′-末端的保守序列(多聚 A 化位点)。给出了每个外显子的长度(以密码子数表示)和每个内含子的长度(以碱基对数表示)。

九、根据 DNA 的多态性可建立家谱(genealogies)

不表达的序列几乎不受选择压力，因而它们的演化比表达的序列快很多，以至于在一个品种内能够积累出显著数目的序列多态性(polymorphism)(变异)。因此，测定一系列的多态性 DNA 序列在一个品种内的不同群体中分布的情况，可以确定这些群体在进化中的亲缘关系。例如，通过人类不同群体的 β 珠蛋白基因簇的 5 个片段中有或无某一限制性酶的酶切位点，即限制性酶切位点长度多态性(RFLPs，见第十四章)技术，已推断出一些不同的人类群体的家谱。通过此研究已建立起一个“家系树”(family tree，图 12-28)，它指明，欧亚人群体彼此之间的亲缘关系比他们与非洲人之间的关系近很多。化石的证据指明，现代人大约在 10 万年前起源于非洲，并很快地遍及全世界。因此，这个家系树提示，所有的欧亚人群都是由一个惊人之小的人群(可能仅有几百人)遗传下来的，这个小的人群大约在 5 万年前离开非洲来到欧亚大陆。

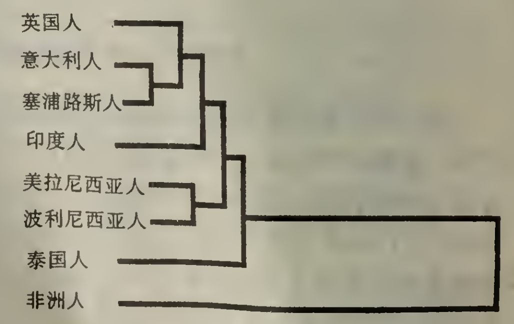


图 12-28　8 个人类群体间血统关系的家族树

这是由他们的 β-珠蛋白基因簇中 5 个限制性位点多态性的分布测定出来的。横坐标指示相关群体间的遗传距离。

十、哺乳动物基因组只有一小部分编码蛋白质

从前面的讨论我们可以把真核DNA分为几类(表12-6),而编码蛋白质的仅为其中的一小部分。以哺乳动物的功能的珠蛋白基因为例来说,α簇和β簇彼此相距很远。α簇长35 kb,β簇长60 kb,平均起来每个基因约为12 kb。然而编码一个约150氨基酸残基的珠蛋白需要DNA的量仅为0.45 kb。可见,这些簇中编码蛋白质的DNA不到其全长的4%。而且,珠蛋白基因和离它最近的编码蛋白质的DNA也相距很远。因此,可能哺乳动物的DNA只有2%是真正编码蛋白质的。可见哺乳动物DNA显然与细菌DNA不同:细菌DNA是非常紧凑的,大量的应用于编码蛋白质。人类基因组约为大肠杆菌基因组的1 000倍,但它的基因数看来只有大肠杆菌的50倍,而不是1 000倍。真核生物的种类很多,它们的基因组结构彼此差异也很大,但可以肯定它们的基因数不是与基因组的大小成比例的。当前分子生物学的一个很大的挑战是阐明那些大量的不编码蛋白质和RNA的DNA的功能。

表12-6　真核DNA的分类

真核DNA的分类
编码蛋白质的基因
单拷贝基因
多拷贝基因
编码RNA的基因
大多数为串联重复
假基因
重复DNA
简单序列DNA
(例如在卫星DNA中)
分散的重复DNA
(包括可移动的遗传成分)
功能不清楚的间隔DNA

第三节　细胞器中的基因组

一、细胞器基因组的发现

在真核细胞中,并非所有的遗传信息都是由其核中染色质DNA编码的。对酵母的遗传研究发现了在线粒体中有与核基因组不同的基因组。1949年,Boris Ephrussi发现,有些面包酵母的突变体是不能进行氧化磷酸化的,这些呼吸缺陷型突变体能够依靠酵解缓慢生长。由于它们形成的菌落很小,所以称之为小菌落(petites)。遗传分析惊异地发现,使小菌落突变分离(segregate)的基因独立于核基因,提示线粒体有其自己的基因组。几年后,发现了线粒体含有DNA。而且一个小菌落株的线粒体DNA的浮力密度(buoyant density)与野生型酵母的不同,在此突变体中其线粒体基因组缺失了很多。随后发现光合真核生物的叶绿体中也含有DNA。线粒体和叶绿体中的DNA都进行复制、转录和翻译。

事实上,早在本世纪初即在植物中已经观察到非孟德尔遗传(non-Mendelian inheritance)现象,这是表明核外有基因组的最早证据。与之有关的体细胞分离(somatic segregation)现象,同样指明有核外基因组的存在。这是因为在非孟德尔遗传中,交配的后代不表现双亲特性的孟德尔分离(Mendelian segregation),反映出其分离特性与减数分裂之间没有联系。而在植物中发现的体细胞分离现象表现为一个植株的某些组织表现出双亲之一的表型,而另些组织则表现出另一个亲本的表型,说明这种分离特性与有丝分裂无关。因此,如果没有核外基因组,则不可能发生这两种现象。而在发现线粒体和叶绿体中有DNA后,这些现象就易于理解了。

非孟德尔遗传的极端情况是单亲遗传(uniparental inheritance),此时只有双亲之一的基因型被遗传,而另一个亲本的基因型则持久地丢失了。非极端的情况是一个亲本的基因型超过

了另一个亲本的基因型。通常的情况是母本的基因型被优先(或单独)遗传,这种现象有时称为母体遗传(maternal inheritance)。重要的是双亲之一对基因型的贡献占优势。这种对双亲基因型的偏倚(bias)是在合子形成时或稍后建立起来的,有多种可能的原因。一种可能是双亲的信息贡献给合子细胞器的不相等,在最极端的情况下是只有双亲之一做贡献;另一种可能是双亲的贡献是相等的,但双亲之一提供的信息未能存活下去;再一种可能是上述两种情况同时都存在。总之,不论是何种原因引起的,这种双亲信息的不等量表现度(representation)是由细胞器遗传物质引起的,它与核遗传信息的情况相反,核遗传信息是由双亲等量获得的。

能够说明此问题的一个例子是在高等动物中发现的一种母性遗传。在受精时合子的线粒体完全是由卵子提供的,而精子则一点也不提供。这样,线粒体的基因完全是由母亲得到的,而且每一代都将雄性的线粒体基因丢弃了。另一个引起单亲遗传的例子是在水藻 *Chlamydomonas reinhardii* 中发现的。它的大小相同的配子(gamete)融合形成合子,这样双亲的叶绿体 DNA 都提供给了合子。但在合子形成之后不久,双亲之一提供的叶绿体DNA 降解了,因而在叶绿体中只剩下了另一个亲本的 DNA。由于这种选择性机制的存在,因而不可能预料双亲之一的基因型占优势的原因。因为即使在双亲提供给合子的细胞浆量不同的情况下,此时可能一个亲本提供的细胞器拷贝数比另一亲本的多,但选择性的机制仍可改变这种情况。

在双亲的细胞器基因组都存活的植物中,在躯体的生长过程中其野生型基因型和突变型基因型发生分离,其结果在一个杂种的(heterozygous)植株中,一些组织有双亲之一的表型;另一些组织则表现为另一亲本的表型,此即为体细胞分离。这是与孟德尔遗传学确定的整个核的基因型的稳定性相反的。体细胞分离发生的原因是因为合子中的细胞器基因是多拷贝的,在体细胞分裂时它们被分配到不同的细胞中,有的细胞分配到一种基因型,而另些细胞分配到另一种基因型的结果。

二、细胞器基因组的一般性质

迄今已鉴定的细胞器基因组大多数是以环状双链 DNA 分子的形式存在。线粒体 DNA 以 mtDNA 表示之,叶绿体 DNA 的代号为 ctDNA。极少数线粒体 DNA 为线型分子,这通常是在低等真核生物中。在每个细胞器中一般有数拷贝的基因组,而每个细胞中又有多个细胞器,所以,在每个细胞中有很多个细胞器基因组。因此,虽然细胞器基因组本身是单一序列的,它不自我重复,但以一个细胞为单位来说,和核中的任何非重复序列相比,它都算是由重复序列组成的。

叶绿体基因组比较大。高等植物的通常为 140 kb 左右,低等真核生物的小于 200 kb。此大小相当于一个大的噬菌体基因组,例如 T4 的约为 165 kb。在高等植物中,每个叶绿体中通常有 20～40 拷贝的基因组,每个细胞中有 20～40 个叶绿体。

不同生物线粒体基因组的大小差别很大。动物细胞的线粒体基因组比较小,人、小鼠和奶牛的由于很小,已经做了全序列分析,都是 16.5 kb 左右。每个细胞中有几百个线粒体,每个线粒体中又有多拷贝的 DNA。果蝇和蛙的线粒体稍大一些,但尚不清楚它们的每个细胞中含有多少个线粒体和每个线粒体中含有几拷贝的基因组。动物线粒体 DNA 的总量比核少,不到核 DNA 的 1%。

酵母线粒体的基因组则大得多。在 *S. cerevisiae* 中,不同株系的大小差别很大,但一般为 84 kb 左右。每个细胞约含 22 个线粒体,每个线粒体约含 4 个基因组。

不同植物线粒体DNA的大小差别很大，但最小的约为100 kb。很多DNA的功能尚未确定，但看来大部分是非编码的。由于其基因组很大，因而难以分离到完整的DNA。但对一些植物进行的限制性酶切图谱分析提示，它们都是单一序列的环状DNA。在此环状DNA中有些短的同源序列，在这些成分之间进行重组（recombination）的结果产生一些小的亚基因组的（subgenomic）环状分子，它们与完整的主（master）基因组共同存在，说明了植物线粒体DNA的复杂性。

细胞器DNA就在该细胞器中进行复制、转录和翻译（与核中不同，核DNA是在核中转录，但到胞浆中进行翻译），所以细胞器中有自己的核糖体和tRNA。细胞器基因组编码其所需要的全部rRNA和tRNA。由于核酸不能穿过细胞器膜，所以细胞器内的各种RNA都是自己合成的。然而转录和翻译所需要的大部或全部蛋白质则是来自胞浆，这些蛋白质是由核基因编码的。表12-7中示出了在线粒体中基因表达所需各种成分的来源。由表中可见，线粒体核糖体是由一个大亚基和一个小亚基组成的，每个亚基含有一个rRNA，它们是由线粒体DNA编码的，而核糖体的蛋白质成分则是由胞浆输入的。然而不同生物的核糖体结构差别很大。线粒体所需一套tRNA的数目比胞浆中的少。例如，人类胞浆中用61种tRNA，而线粒体中仅用22种，有些线粒体tRNA阅读4种密码子。虽然线粒体的蛋白质合成装置的所有蛋白质成分都是由核基因编码的，但这些蛋白质成分是线粒体独有的，与胞浆中的并不相同。因而细胞器中的蛋白质合成有与胞浆中不同的特征。

表12-7　线粒体中基因表达所需成分的来源

成　分	在线粒体中合成的	在胞浆中合成的
模　板	~10mRNAs	RNA聚合酶
核糖体	2rRNAs	~75蛋白质
衔接子	22tRNAs	22AA-tRNA合成酶

线粒体和叶绿体都是与能量转换有关的细胞器，因而它们的基因组除了编码rRNA和tRNA外，主要编码它们所需要的与能量转换有关的蛋白质。这些基因只在细胞器中转录和翻译，并且细胞器只表达它自己的基因。然而，细胞器只产生它所需要的蛋白质的一部分；另一部分则来自胞浆，是由核基因编码的。细胞器用这两部分蛋白质构建成能量转换体系。以酵母线粒体为例，它的呼吸链成分的来源列于表12-8。由表中可见，酵母呼吸链中的几个蛋白复合体的成分只有一小部分是线粒体自己合成的，其余大部分由胞浆中运入，然后共同组装成复合体。例如，ATP酶由两个单位组成：一个是由3个亚基组成的膜因子，这3个亚基是由线粒体基因组编码的；另一个是F1ATP酶，它的5个亚基则是在胞浆中合成的。细胞色素c氧化酶的亚基也是来自两个方面。而细胞色素bc1复合体只有一个亚基是由线粒体合成的，其它6个亚基都是来自胞浆。

表12-8　酵母线粒体呼吸链中蛋白复合体的成分来源

蛋白复合体	分子量（kD）	线粒体中合成的亚基	胞浆中合成的亚基
寡霉素敏感ATP酶	340	ATP酶6，8，9（F0膜因子）	ATP酶1，2，3，4，7（F1ATP酶）
细胞色素c氧化酶	137	CO 1，2，3	CO 4，5，6，7
细胞色素bc1复合体	160	细胞色素b脱辅蛋白质	6亚基

细胞器中的条件与核中的不同，因而细胞器的 DNA 以其特有的速度进化。如果遗传是单亲的，则没有双亲基因组间的重组。并且即使细胞器基因组是由双亲遗传下来的，通常也不发生重组。由于复制细胞器 DNA 的 DNA 聚合酶与核中的不同，因而在复制中的错误率也就不同。再者，细胞器的 DNA 修复体系以及其它影响序列忠实性(fidelity)的机制也与核中的不同。由于这些原因，已知哺乳动物线粒体 DNA 聚累突变的速度比核中快。而在植物中其线粒体聚累突变的速度比核中慢，叶绿体的速度则在线粒体与核的速度之间。

母性遗传的后果之一是线粒体 DNA 序列对育种群(breeding population)规模的降低比核 DNA 更敏感。这意味着分析线粒体 DNA 序列间的差异比分析核 DNA 的差异更易于区分个体。因此，在育种中为了用遗传标志(例如，用限制性片段长度多态性技术)区分群体时，当前大家更趋向于分析线粒体 DNA，而不是核 DNA。

由于不同生物的细胞器基因组差别很大，目前对大多数生物的细胞器基因组尚不清楚，下面仅举几个比较清楚的例子，以便对细胞器基因组有些具体的概念。

三、酵母线粒体基因组

面包酵母(*S. cerevisiae*)的线粒体基因组约为 84 kb，哺乳动物的约为 16 kb，酵母线粒体基因组的大小约为哺乳动物的 5 倍。这一巨大差异本身提醒我们，虽然酵母线粒体的功能与哺乳动物的一样，但它们的基因组构(gene organization)必定有大的差异。从与线粒体功能有关的酶来说，在酵母线粒体内合成的数目与哺乳动物是相似的，那么酵母线粒体中额外的遗传物质的功能究竟是什么?

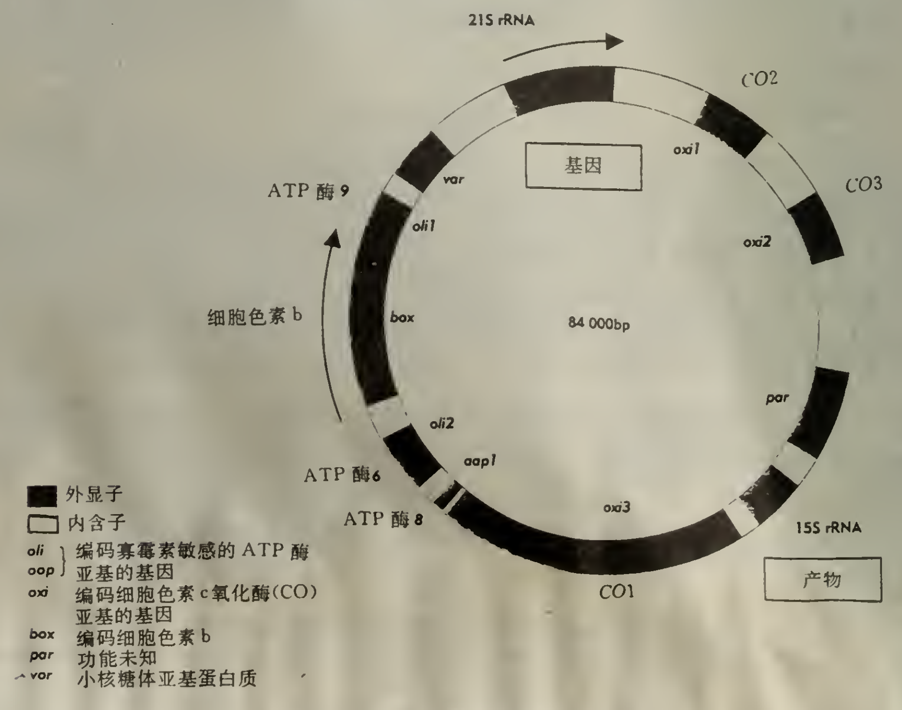


图 12-29　面包酵母的线粒体基因组

tRNA 的基因未列出，箭头指明转录的方向。

图 12-29 是酵母线粒体的基因图谱，指明了其主要的 RNA 和蛋白质产物。但没有指出其约 22 个 tRNA 分子，因为它们还没有被完全定位。此基因组的一个罕见的特点是它的两个 rRNA 基因是分开的。其 15 S rRNA 基因是不割裂的，并且与 21 S rRNA 基因相距约 24 kb。在面包酵母的某些株系中 21 S rRNA 基因有一个内含子(如图中所示)，但在另一些株系中则是不割裂的。

两个最突出的基因座是 *box*(编码细胞色素 b)和 *oxi*3(编码细胞色素氧化酶的亚基 1)，它们都是割裂基因。这两个基因加在一起几乎与哺乳动物线粒体的整个基因组一般长！在这些基因中的许多长内含子和它们前面的外显子合在一起有开放读框，有证据表明，至少在有些情况下内含子是被翻译的。这就使酵母线粒体增加了一些蛋白质，推测这些蛋白质的合成量是低的。

其余的基因看来都是非割裂的。其中有编码细胞色素氧化酶的另外两个亚基的；编码 ATP 酶亚基的，以及 *var* 是编码线粒体核糖体蛋白质的等。酵母线粒体基因组有 24%左右是由短的富含 A═T 碱基对的序列所组成，它们大概没有任何编码功能。尽管如此，仍然还有相当大量的遗传物质的功能需要鉴定，并且，如果在图谱的空白区域不能进一步找到一些基因，那将是令人奇怪的事。然而，即使再找到一些基因，估计酵母线粒体基因的总数也不会超过 25 个左右。

四、哺乳动物的线粒体基因组

动物线粒体 DNA 的一个突出特点是它非常紧凑。虽然不同门(phyla)的动物线粒体基因组构(gene organization)的细节很不相同，但基本情况是相同的。它们的基因组都比较小，编码的数目也比较少。知道得最清楚的是哺乳动物的线粒体基因组，其中有些动物的已做了全序列分析。它们的基因组构非常紧凑，没有内含子；有些基因实际上是重叠的；并且几乎每一个碱基对都属于一个基因。只有 D 环(D loop)是例外，它是 DNA 复制的起始区。在人类线粒体基因组的 16 569 bp 中，可算成是基因之间的区域的不超过 87 bp。

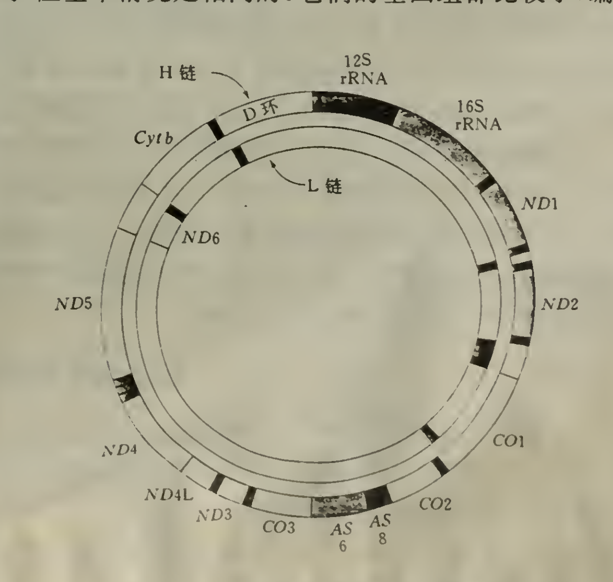


图12-30　人类线粒体基因组的图谱，指明了由两条链(H 和 L)编码的基因

ND 基因编码 NADH-Q 还原酶(也叫 NADH 脱氢酶或复合体 1)的亚基；*CO* 是细胞色素氧化酶的基因；*Cytb* 是细胞色素还原酶 *b* 成分的基因；*AS* 是 ATP 合成酶的基因；■为 tRNA 的基因。

人类和鼠的线粒体基因组的全部核苷酸序列表现出高度的同源性。图 12-30 是人类线粒体基因组的概要。它是个环状 DNA 双螺旋，含有 16 569 bp。此基因组编码 13 个蛋白质，22 个 tRNA 和 2 个 rRNA。此 DNA 的蛋白质编码容量约 60%用于编码 NADH-Q 还原酶的 7 个亚基，此酶是线粒体内膜上 3 个质子泵复合

体(proton-pumping complexe)中的第一个。此DNA分子还编码细胞色素还原酶的一个亚基，细胞色素氧化酶的3个亚基，和ATP合成酶的2个亚基。其中L链(L strand，两条链中密度较小者)编码一种蛋白质和14种tRNA，其余所有的蛋白质和RNA都由H链(H strand，两条链中密度较大者)编码。

在许多情况下是基因之间完全没有间隔。一个基因的最后一个碱基与下一个基因的第一个碱基相邻。有些情况是有重叠的，最常见的是仅重叠一个碱基，即一个基因的最后一个碱基是下一个基因的第一个碱基。有5个读框缺乏终止密码子，它们的末尾是U或UA，因而是通过转录物的加多聚A尾巴(线粒体mRNA在其3′末端加一个短的多聚A序列)而补充产生终止密码子UAA。有3种情况是其终止密码子为AGA或AGG，它们通常是用精氨酸的密码子。相当于所有蛋白质编码区的mRNA已被鉴定。每个mRNA的起始密码子都离转录起点6个碱基以内。所用起始密码子可以是AUG，AUA或AUU。人类线粒体mRNA实际上不存在5′和3′的非翻译区。人类线粒体有4个密码子的意义是不标准的：正常编码精氨酸的AGA，AGG为终止密码子；AUA编码甲硫氨酸而不是异亮氨酸；UGA编码色氨酸而不是终止信号。

除了上述基因组的利用非常经济和有些密码子与核中有所不同外，人类线粒体基因组在基因的表达方面还有个明显的特点，其表现为：①几乎所有基因的转录方向都是顺时针的；②除了有一处有两个基因(ATP酶6和*CO*3)相邻外，所有其它基因都是用至少一个tRNA基因所隔开；③用tRNA基因将rRNA和蛋白质编码区断开之处并没有像真核的核中或细菌中那样的启动子，而是仅在D环区有一个顺时针转录的启动子。这些特点提示，转录只是从12 S rRNA基因前面的tRNA基因之前开始，一直把整个环转录完，到D环区终止。因此，由H链产生一条转录物，L链产生另一条转录物。这些初始转录物在tRNA的两端断开，形成单基因的mRNA，只有ATP酶6和*CO*3是例外。因此，线粒体DNA与细菌的一个操纵子很相似，它由单一的启动子区开始转录，转录出一条初始转录物，然后经过加工每个tRNA，rRNA和mRNA释放出来，可见在其基因表达中加工是非常重要的。

L链编码的tRNA基因是怎样表达的?其机制大概与上述类似。L链产生一个大的转录产物，在其中有tRNA序列和mRNA序列，然后经过降解，把这些序列保存下来，而其余的则降解掉。L链也是由D环区开始转录(逆时针)。可以提取到很大的L链转录产物，说明它是在转录后经过一定时间才加工的。而H链的大的转录产物则未能获得，说明它是在转录出每个tR-NA时(转录尚未终止)便立即加工的。

五、叶绿体基因组

叶绿体的基因列于表12-9。叶绿体基因组的情况与线粒体的类似，只是比线粒体的基因多一些。叶绿体基因组编码其合成蛋白质所需要的所有rRNA和tRNA。在它的核糖体中除了两个主要的rRNA外，还含有两个小的rRNA。和线粒体类似，叶绿体tRNA的种类也比胞浆中的少。叶绿体基因组编码约50种蛋白质，其中包括RNA聚合酶和一些核糖体蛋白质。叶绿体的基因也是在叶绿体中转录和翻译的。

不同高等植物的叶绿体基因组长度不同，但都有个特点，即都有个长的序列，其长度因不同植物而不同，在10～20 kb范围内，它以两个相同的拷贝存在，并成为反向重复(图12-31)。在此反向重复之内的基因是每个基因组两拷贝，并且含有rRNA基因，两个重复序列之间的重组，可使它们之间的短的单拷贝序列(short single copy sequence，SSC) 相对于长的单拷贝

序列(long single copy sequence，LSC)来说发生逆转。

已经测定了欧龙亚草(*Liverwort*，一种苔藓)、烟草和水稻叶绿体 DNA 的全序列，水稻叶绿体基因组图谱见图 12-32。虽然它们的总长度相差很大(欧龙亚草的约为 121 kb，烟草的约为 155 kb)，但基因的组构相似，基因的总数几乎相同。基因产物的大多数是类囊体(thylakoid)膜的成分，或与氧化还原反应有关的成分．如表 12-9 中所示。

表 12-9　叶绿体基因组编码 4 个 rRNAs，30 个 tRNAs 和 50 种左右蛋白质

基因	类型
编码 RNA	
16 S rRNA	1
23 S rRNA	1
4.5 S rRNA	1
5 S rRNA	1
tRNA	30
基因表达所需的	
核糖体蛋白质	19
RNA 聚合酶	3
其它	2
类囊体膜上的	
光系统 Ⅰ	2
光系统 Ⅱ	7
细胞色素 b/f	3
H^+-ATP 酶	6
其它	
NADH 脱氢酶	6
铁氧还蛋白	3
RuBP 羧化酶	1
未查明的	29
总计	115

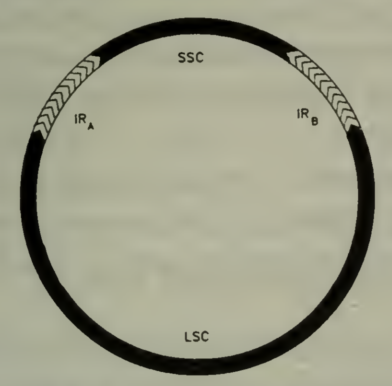


图 12-31　叶绿体基因组是一个环，它被两个反向重复序列分开为一个短的单拷贝区和一个长的单拷贝区

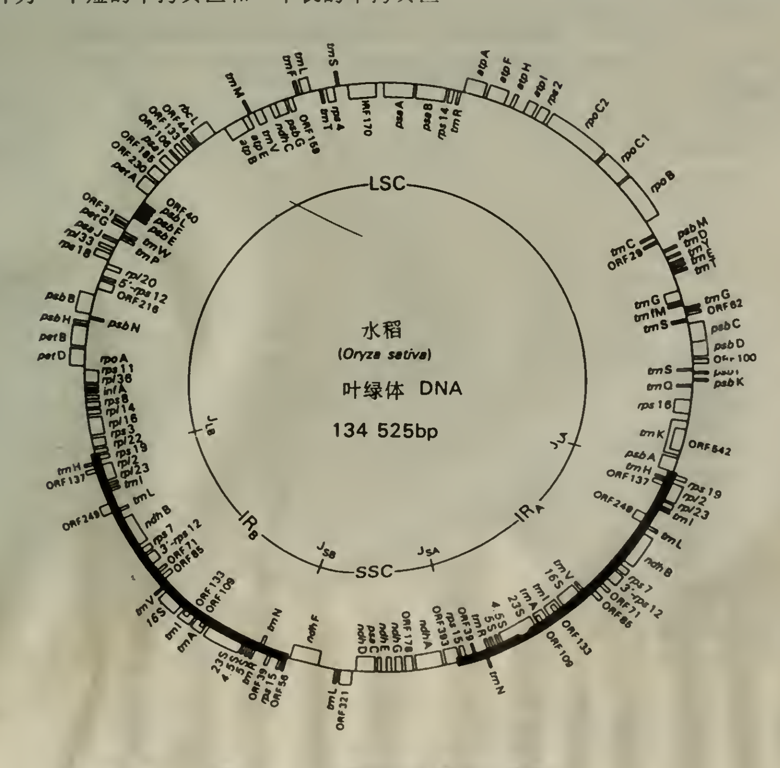


图 12-32　水稻叶绿体基因组图谱

杉浦昌弘等 1986，1988.

表 12-9　欧龙亚草、烟草及水稻叶绿体 DNA 全序列及其组构

	欧龙亚草(bp)	烟草(bp)	水稻(bp)
反向重复序列(IR)	10 058	25 339	20 798
短单拷贝区(SSC)	19 813	18 482	12 335
长单拷贝区(LSC)	81 095	86 684	80 592
总基因组	121 024	155 844	134 525

叶绿体的功能是进行光合作用，它的许多基因是编码类囊体膜的复合体中的蛋白质。表 12-10 中列出了烟草叶绿体基因组中所含与光合作用有关的基因。光合作用可分为暗反应和光反应两部分。在暗反应的酶系中只有核酮糖二磷酸羧化酶-加氧酶(rubisco)的大亚基是由叶绿体 DNA 编码的。在光反应中有 5 组类囊体膜上的复合体，即 ATP 酶、光系统Ⅰ、光系统Ⅱ、细胞色素 b/f 复合体和集光叶绿体蛋白复合体。

从表中可以看出，虽然有些复合体和线粒体的情况一样，即复合体的一些亚基由叶绿体基因组编码，而另些亚基则由核基因组编码，但叶绿体有些复合体的成分是由一种基因组编码的。例如，光系统Ⅰ、光系统Ⅱ的亚基完全由叶绿体基因组编码，而集光叶绿体蛋白复合体的亚基则完全由核基因组编码。

已被鉴定出来的叶绿体基因的情况如下：45 个编码 RNA 的基因；27 个基因编码与基因表达有关的蛋白质；18 个基因编码类囊体膜的蛋白质；另外有 10 个基因编码与电子传递功能有关的蛋白质。还有约 30 个开放读框的产物尚待鉴定。

表 12-10　烟草叶绿体 DNA 所含编码的光合作用中的基因

与光合作用有关的成分		编码的基因名称	氨基酸残基数	分子量(kD)	与菠菜同源性(%)
Rubisco	大亚基	*rbc L*	477	53	92
H^+-ATPase	α 亚基	*atP A*	507	55	87*
	β 亚基	*B*	498	54	96
	ε 亚基	*E*	133	15	87
	Ⅰ 亚基	*F*	184	19	78(有 695 bp 内含子)
	Ⅲ 亚基	*H*	81	8.0	99
	a 亚基	*I*	247	27	92♀
光合系统Ⅰ	P700 前蛋白	*psa A*	750	83	97#
	A2 亚基	*B*	734	82	96#
光合系统Ⅱ	P32 蛋白	*psb A*	353	39	100
	P680 前蛋白	*B*	508	56	96
	44 kD 蛋白质	*C*	473	52	98
	D2 蛋白质	*D*	353	40	98
	细胞色素 b559	*E*	83	9.4	99
	10 kD 磷酸化蛋白	*F*	73	7.8	
	4 kD 蛋白质(ORF39?)	(*G*)	39	4.5	97
电子传递系统	细胞色素 f	*pet A*	320	35	91
	细胞色素 b6	*B*	215	24	97(有 759 bp 内含子)
	亚基 4	*D*	139	15	98

* 和小麦比较；# 和玉米比较，♀：和豌豆比较。

叶绿体的蛋白质合成装置表现出与细菌的装置相似。许多核糖体蛋白质(r 蛋白质)与大肠杆菌的是同源的，而且核糖体蛋白质的一些基因的成簇存在也与大肠杆菌中的相应的簇类

似。RNA 聚合酶的 α，β 和 β' 亚基的基因与大肠杆菌核心酶(core enzyme)的基因同源。然而其 β' 亚基的基因有个内含子。割裂基因的存在提示叶绿体来源于丢失其内含子之前的原核生物。叶绿体的内含子有两类：①tRNA 基因的内含子通常(不是必定)位于反密码子环中，类似于酵母核中的 tRNA 基因；②编码蛋白质的基因中的内含子类似于线粒体基因的内含子。

第四节　mRNA 前体的剪接

由于真核生物，特别是高等生物的编码蛋白质的基因绝大多数是割裂的，并且已经证明在转录时其内含子与外显子一起转录，因而最初的转录产物是含有内含子的未成熟的 mRNA，称为 mRNA 前体。随后必须把内含子删除掉，将外显子联结在一起，才能形成成熟的 mRNA，作为合成蛋白质的模板。这一删除内含子的过程称为 RNA 的剪接(splicing)，已知它是在核中进行的。此外，已知有些 rRNA 和 tRNA 的初始转录物也有内含子，它们也需要经过剪接才能成为功能的 rRNA 和 tRNA。研究 rRNA 前体的剪接导致了核酶(ribozyme)的发现。

一、mRNA 前体的剪接位点是由内含子末端的序列规定的

mRNA 前体中的内含子必须精确地剪接掉，因为在剪接位点处多剪接掉或少剪接掉一个核苷酸，就使得剪接处 3′一侧的所有密码子都改变了，从而合成出一个完全不同的氨基酸序列。那么，剪接的机械是怎样做到高度忠实的对其底物进行加工的？已分析了几百个剪接前的 RNA 转录物中内含子-外显子交界处的碱基序列(表 12-11)，从而解答了此问题。

表 12-11　含有内含子的转录物中剪接区的碱基序列

基因区域	外显子	内含子	外显子
卵清蛋白内含子 2	UAAG	GUGAGC ～～ UUACAG	GUUG
卵清蛋白内含子 3	UCAG	GUACAG ～～ AUUCAG	UCUG
β-珠蛋白内含子 1	GCAG	GUUGGU ～～ CCUUAG	GCUG
β-珠蛋白内含子 2	CAGG	GUGAGU ～～ CCACAG	UCUC
免疫球蛋白 λ_1 内含子 1	UCAG	GUCAGC ～～ UUGCAG	GGGC
SV40 病毒早期 T-抗原	UAAG	GUAAAU ～～ UUUUAG	AUUC

在测定这些序列之初曾经设想，在内含子两端的碱基序列可能是彼此互补的。由这些互补序列通过碱基配对而形成发夹结构，从而使两个外显子靠拢在一起，这样在酶的作用下，便易于将内含子剪除，把外显子联结在一起。然而分析的结果和设想的完全不同，没有发现有互补序列。但是在分析了从酵母到哺乳动物这样广泛范围内的真核生物的这些序列之后，发现它们有个共同的结构基元(structural motif)：即所有内含子的序列都是以 GU 开始，以 AG 告终。脊椎动物内含子 5′-末端的共有序列(consensus sequence)为 AGGUAAGU(图 12-33)。在内含子 3′-末端的共有序列是在一段 10 个嘧啶(U 或 C)之后有个任意碱基，然后是个 C，最后是不变的 AG。内含子还有一个重要的内部位点，它位于 3′剪接位点上游 20 和 50 个核苷酸之间，

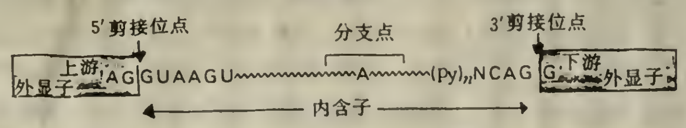


图 12-33　剪接信号，指明了 5′剪接位点和 3′剪接位点的共有序列

叫做分支点(branch site),叫做分支点的原因在随后讨论剪接过程时就明白了。在酵母中分支点的序列几乎都是UACUAAC,但在哺乳动物中已发现了多种不同的序列。

在内含子中,除了5′和3′剪接位点以及分支点的序列之外,其它部分对决定在何处发生剪接看来并不重要。内含子的长度范围为50～10 000个核苷酸。一个内含子的大量序列被缺失掉并不改变其剪接位点或效率。同样,在基因的内含子中插入大片段的DNA也不影响其剪接。而且,用重组DNA的方法将一个内含子的5′端一段序列和另一个十分不同的内含子的3′端一段序列联结在一起,形成一个嵌合的内含子,只要其剪接位点和分支点没有改变,照样发生正确的剪接。与之相反,在这3个关键区域的任一个区域发生突变时,则可引起异常的剪接。在离剪接位点的远处发生突变时也能引起不正常的剪接,但这种情况只有当突变后在新的位置产生了一个5′或3′的剪接共有序列时才会发生。

地中海贫血症(thalassemia)是一类遗传性贫血症,其特点是血红蛋白的合成有缺陷。在这类贫血症中有一些是由于异常剪接引起的。其中有一个病例是在其β珠蛋白基因的第一个内含子中发生了突变,此突变发生在正常3′剪接位点上游离位点20个核苷酸处,正常此处为G,病人的突变成了A(图12-34),从而产生了一个新的3′剪接位点。由于用此新的剪接位点剪接的结果,使产生的mRNA中的一系列密码子不正常了。在新的剪接位点剪接后的第7个密码子是蛋白质合成的终止信号,因此合成出来的是个不正常的未成熟蛋白质,所以就产生了贫血病。

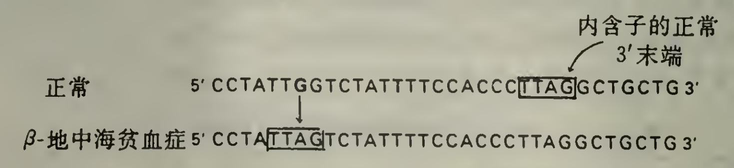


图12-34 由于β珠蛋白基因的一个内含子发生了一个碱基突变(G变为A)引起的地中海贫血病
此突变大概是产生了一个与正常3′剪接位点相类似的新的3′剪接位点,此位点在正常位点的上游。引自R. A. Spritz等。

二、剪接体催化的剪接

首先讨论在真核生物中最常见的一种剪接方式,即剪接体(spliceosome)催化的剪接方式。

(一)在mRNA前体的剪接中形成套索形中产物

在一个mRNA的前体中,一个外显子的末端是怎样与其相邻的另一个外显子的前端联结在一起的?剪接的第一步是上游外显子(外显子1)与内含子5′末端之间磷酸二酯键的裂解(图12-35)。在此反应中攻击的基团是分支点中的一个腺苷酸残基的2′-OH。通过此反应,在此腺苷酸残基与内含子的5′-末端磷酸基之间形成了一个2′,5′-磷酸二酯键。注意,此腺苷酸残基仍然用正常的3′,5′-磷酸二酯键联结在另外2个核苷酸上(图12-36)。因而在此位点形成了1个分支,并形成了1个套索形的(lariat)中产物。然后,外显子1的3′-OH末端攻击内含子和外显子2之间的磷酸二酯键,其结果是外显子1和外显子2联结在一起,并将内含子以套索形释放出去。值得注意的是,剪接是由2个转酯反应(transesterification reaction)完成的,而不是由水解后再联结完成的。第一个反应在外显子1的末端产生一个3′-OH,而在第二个反应中此3′-OH与外显子2的5′-磷酸基联结在一起。在这些反应步骤中,磷酸二酯键的总数始终没有改变。第一个反应的产物一直由剪接体抓持在一起,直到两个外显子联结起来之后为止。剪接体是由多个核蛋白聚集而成的,它能识别mRNA前体的5′剪接位点、3′剪接位点和分支点,下面讨论之。

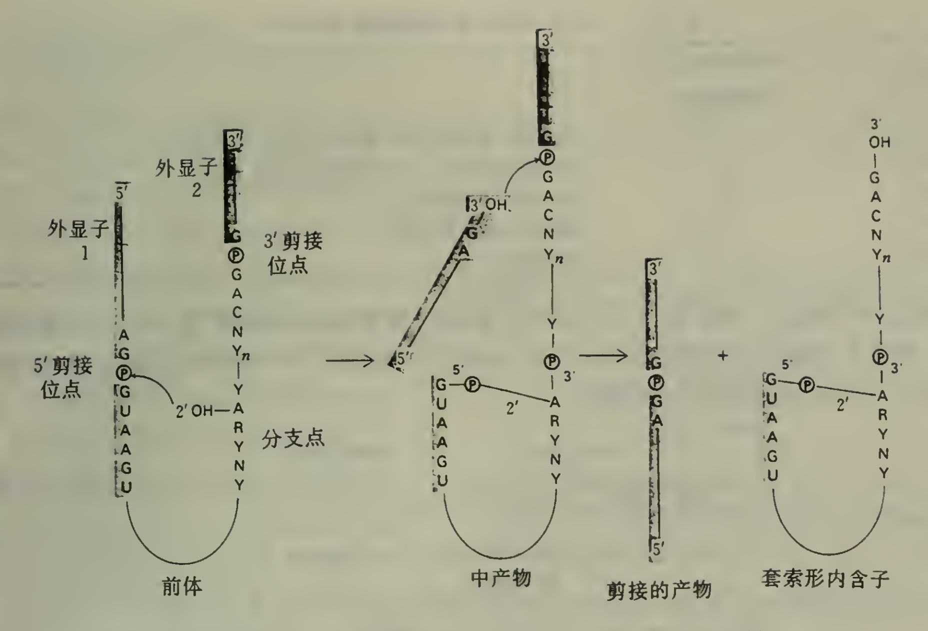


图 12-35　在真核生物的细胞核中 mRNA 剪接的机制

Y 代表一个嘌呤核苷酸，R 代表一个嘧啶核苷酸，N 代表任何核苷酸。5′剪接位点受到分支点腺苷 2′-OH 的攻击，然后新生成的上游外显子的 3′-OH 攻击 3′剪接位点。两个外显子联结在一起了，内含子以套索的形式释放出去。引自 P. A. Sharp，Cell. 1985. 2：3 980.

（二）剪接体中的核内小 RNA（snRNA）催化 mRNA 前体的剪接

在细胞核和胞质溶胶中都含有多种不到 300 个核苷酸的小 RNA 分子，它们被称为核内小 RNA（small nuclear RNA，snRNA）和胞质内小 RNA（small cytoplasmic RNA，scRNA）。这些 RNA 分子和特异的蛋白质结合在一起形成复合体，这些复合体分别叫做核内小核糖核蛋白颗粒（small nuclear ribonucleoprotein particles，snRNPs）和胞质内小核糖核蛋白颗粒（small cytosolic ribonucleoprotein particles，scRNPs）。研究者们也常常把它们叫做“snurps”和“scurps”。剪接小体是一类大的复合体（60 s），它们是由 mRNA 前体和 snRNPs（表 12-12）动态的聚集在一起而形成的。由于发现特异于这些核糖核蛋白的抗体能够封阻剪接作用，从而揭示出这些 snRNPs 在剪接中起着关键的作用。这些特异套的抗体是由全身性红斑狼疮（systemic lupus erythematosus）病人中获得的。这是一种自身免疫（autoimmune）的疾病，其特点是产生抗许多细胞核成分的抗体。这些抗体与核成分相结合，引起了关节炎和肾损伤。

图12-36　在剪接过程中形成的索形中产物中分支处的结构

此腺苷酸残基以磷酸二酯键与 3 个核苷酸相联。图中腺苷酸与鸟苷酸之间的是新形成的 2′→5′磷酸二酯键。

表 12-12　参与 mRNA 前体剪接的 snRNPs

snRNP	snRNP 的大小 （核苷酸数）	作用
U1	165	结合在 5′剪接位点上，然后结合在 3′剪接位点上
U2	185	结合在分支点上，并形成催化中心的一部分
U5	116	结合在 5′剪接位点上
U4	145	遮蔽 U6 的催化活性
U6	106	催化剪接

在哺乳动物细胞中，剪接是以 U1 snRNP 识别 5′剪接位点开始的（图 12-37）。事实是，在 U1 RNA 中有个与此位点互补的序列，当 U1 snRNP 结合在一个 mRNA 前体上时，保护了 5′剪接位点处一个 15-核苷酸区不被消化。

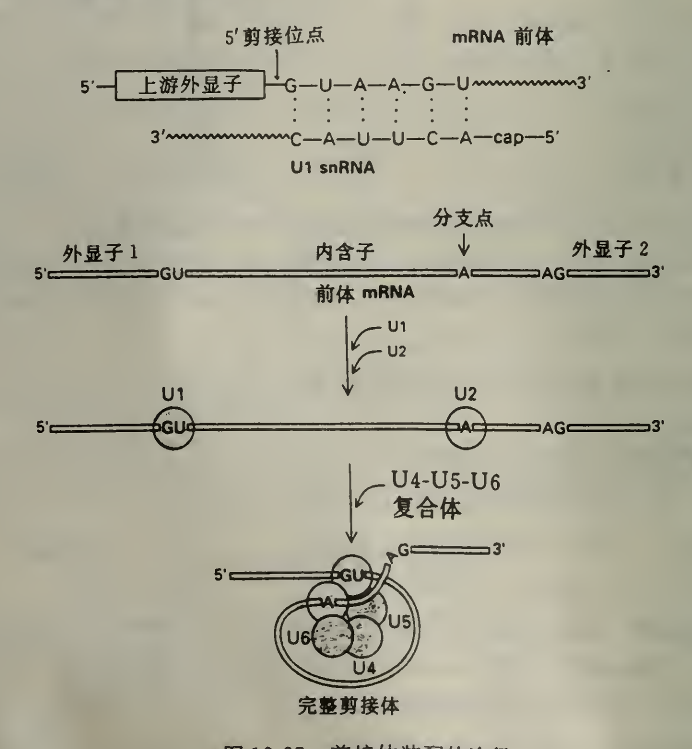


图 12-37　剪接体装配的途径

U1 结合在 5′剪接位点上，U2 结合在分支点处。一个事先形成的 U4-U5-U6 复合体最后联结在其上形成完整的剪接体。

然后 U2 snRNP 结合在内含子的分支点上。随后由于 U1 和 U2 结合起来的结果，把内含子的 5′端和 3′端带到一起，使 U1 也能与 3′剪接位点配对。此 U1，U2 和 mRNA 前体形成的复合体与事先组装好的 U4-U5-U6 复合体联结在一起，形成一个完整的剪接体。

已有的证据提示剪接的途径如下：首先，U5 与 5′剪接位点处的外显子序列相互作用。下一步是 ATP 推动的 U1 与 5′剪接位点之间的碱基配对的解开。U1 离开剪接体，U5 扩展其与 mRNA 前体的碱基配对至包括 5′剪接位点在内。再后，U4 和 U6 之间的碱基配对折开，从而释放出 U6 的催化活性，开始了剪接反应。U4 是个抑制剂，在对准特异的剪接位点之前，它一直把 U6 遮蔽起来。

U6 从 U4 释放出来后，它就与内含子 5′剪接位点的保守序列进行配对。此外，U6 还与 U2 配对，而 U2 又与内含子的分支点配对。U2 和 U6 snRNAs 大概是构成了剪接体的催化中心(图 12-38)。然后分支点的 2′-羟基攻击 5′剪接位点，释出外显子 1，新生成的外显子 1 的 3′-羟基再攻击 3′剪接位点，从而把两个外显子联结在一起。联结在剪切下来的套索形内含子上的 U2，U5 和 U6 最后被释放出来，剪接反应就完成了。此过程有两个特点值得注意：①RNA 分子在对准剪接位点和在进行催化中起着关键的作用；②在 ATP 的推动下，蛋白质使 RNA 的双链中产物解开，并且引起核糖核蛋白由 mRNA 前体及其产物上释放出来。

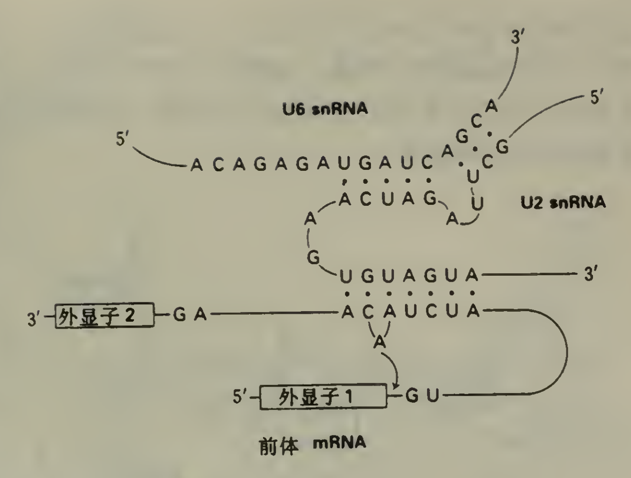


图12-38 剪接体的催化中心是由 U2 snRNA 和 U6 snRNA 形成的，它们之间形成了碱基配对

引自 H. D. Madhani 和 C. Guthrie. Cell. 1992. 71：803.

三、自我剪接的 RNA：催化活性的 RNA 的发现

和原核的一样，真核生物的核糖体 RNA 分子也是由初始转录物经过裂解后生成的。在四膜虫(*Tetrahymena*，一种有纤毛的原生动物)中，裂解后生成的一个 6.4 kb 的前体中要除去 1 个 414 核苷酸的内含子以生成成熟的 26 S rRNA 分子(图 12-39)。Thomas Cech 及其同事在研究此剪接反应时得到了一些非常意外的发现。他们开始研究时的目的是想鉴定出剪接作用所需要的蛋白质，做法是在制得此 RNA 前体的纯品后，向其中加入细胞抽提物，以便观察细胞抽提物中的那些蛋白质能促进剪接。然而他们惊异地发现，一个仅含有纯化的前体和核苷三磷酸，而显然没有蛋白质的对照样品中居然发生了剪接作用。由于剪接作用是个有共价键的裂解和生成反应的过程，这种过程是必须有酶参与催化的，而在任何酶的化学本质都是蛋白质已经成为固定概念的当时，这种没有蛋白质的参与即发生了剪接作用的现象是没有人能够相信的。因而怀疑，是不是在制备的前体 RNA 中没有将蛋白质除尽？而残留的一点点蛋白质正是催化剪接作用所必需的？为了解答这个问题，他们用重组 DNA 的方法制备与此 6.4 kb 相应的 DNA。这是用大肠杆菌作为寄主细胞制备的，而大肠杆菌细胞中原来是没有此 RNA 分子的。在获得纯的编码此前体 RNA 的 DNA 后，再用体外转录的方法制备出合成的此 RNA，并观察其剪接反应。其结果与上述相同，即在有核苷酸存在的条件下，此 RNA 自我剪接了，并且精确地将其 414 核苷酸的内含子剪接掉了。这一卓越的实验表明，一个 RNA 分子能够具有高度特异的催化活性，并能自我剪接。

前已述及，在自我剪接反应中需要有核苷酸的参与。在原来寻找起催化作用的蛋白质时在反应混合物中加入 ATP 或 GTP，是因为曾经设想剪接反应需要由 ATP 或 GTP 提供能量。后来在研究自我剪接的反应时发现，它不需要 ATP 或 GTP 提供能量，但此反应需要 GTP 作为辅助因子。而且不一定非 GTP 不可，鸟苷、GMP，GDP 或 GTP 中任何一种都可起辅助因子的作用，我们用 G 代表这些化合物中的任何一种。G 的作用不是提供能量，而是作为一个攻击基

团并暂时掺合到RNA中(图12-39)。G结合在RNA上之后,它攻击5′-剪接位点,并与内含子的5′-末端形成磷酸二酯键。此转酯反应使上游外显子产生一个3′-OH,然后此新生成的上游外显子的3′-OH攻击3′-剪接位点。此第二个转酯反应把两个外显子联结在一起,并将414核苷酸的内含子释放出去。

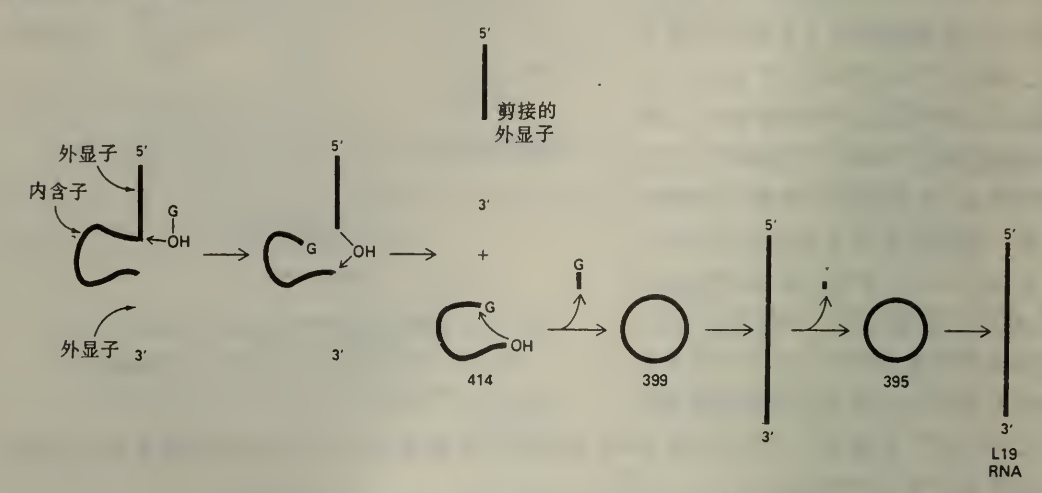


图12-39　一个四膜虫的核糖体RNA前体的自我剪接作用

剪接过程中的第一个剪接反应是将一个414核苷酸的内含子释出,在此反应中鸟苷或者一个鸟苷酸(用G表示)是个辅助因子,随后此内含子又自我剪接两次,产生一个线型RNA,它比原来的内含子少了19个核苷酸。此L19 RNA是有催化活性的。

随后还发生两轮自我剪接。第一轮剪接是生成的内含子的3′-OH攻击其靠近5′-末端的一个磷酸二酯键,结果内含子形成环状,并释出一个含有G的15核苷酸片段,G是在前述的剪接过程中掺合进来的。生成的399核苷酸环打开成为线型分子,并进行第二轮剪接,其结果是释出一个4核苷酸片段并形成环。此环再打开形成一个线型RNA,此RNA叫做L19 IVS(linear minus 19 intervening sequence,线型的缺失19个核苷酸的间插序列),我们称之为L19 RNA。

自我剪接依赖于此rRNA前体分子结构的完整。自我剪接对内含子有许多要求。与许多其它RNA一样,此RNA分子具有折叠结构,它是由许多双螺旋的茎(stem)和环(loop)构成的(图12-40)。在此折叠的RNA中除了含有较强的AU和GC碱基对外,还含有较弱的GU碱基对。这些碱基对中两个碱基之间的亲和力大致为:GU∶AU∶GC=1∶100∶1 000。在加入变性剂,如二甲基甲酰胺等,使rRNA前体的二级和三级结构遭到破坏时,剪接作用便被封阻了。而且,G的结合是可饱和的(K_m值为32 μm),并且能被竞争性地抑制。这些发现强有力地提示,在此前体中含有一个特异结合G的袋(pocket)(图12-41)。

碱基序列分析的结果提示,5′剪接位点与催化位点是以碱基配对的方式相结合的,即在上游外显子中的一个富含嘧啶区(CUCUCU)与内含子中的一个富含嘌呤的指导序列(guide sequence,GGGAGG)相配对(图12-42)。内含子将辅助因子鸟苷和5′剪接位点聚拢在一起,使得G的3′-OH正好能亲核攻击此剪接位点的磷原子。然后内含子的另一部分将下游外显子抓持在一定位置,使新生成的上游外显子的3′-OH正好攻击它。其结果是在两个外显子之间形成一个磷酸二酯键,同时将内含子以线型分子的形式释放出来。在此rRNA前体中自我催化的键的生成和裂解是高度立体特异性的(stereospecific),这和化学本质为蛋白质的酶所催化的反应是一样的。

图 12-40　四膜虫的一个核糖体 RNA 前体的二级结构的一部分

引自 T. Cech. RNA as an Enzyme. Copyright © 1986 by Scientific American, Inc.

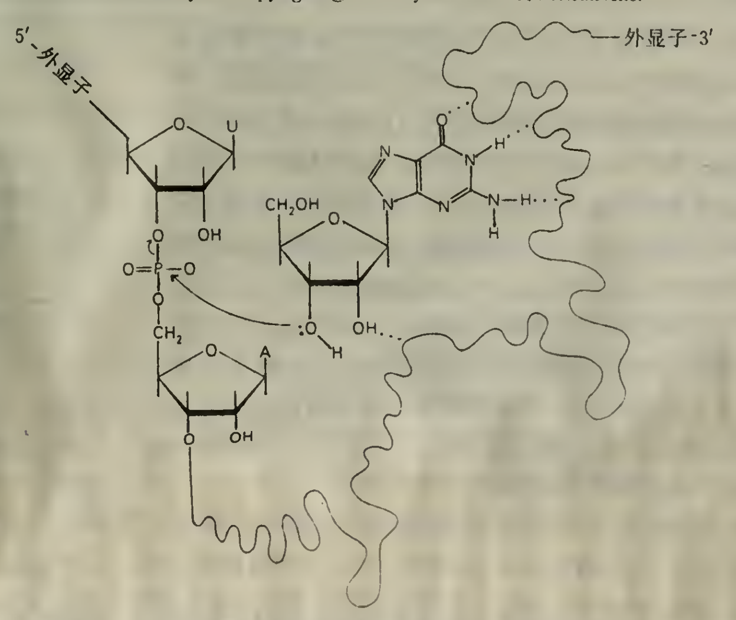


图 12-41　设想的一个 RNA 酶中含有的活性中心处结合鸟苷的模式图

引自 B. L. Bass and T. R. Cech. Nature. 1984. 308: 820.

在线型的内含子分子释出后，它的富含嘌呤的指导序列再与内含子中的另一个富含嘧啶的序列相结合，从而进行再一次剪接，释出一个15核苷酸的片段，并形成环状。此环状RNA打开，另一个富含嘧啶的序列再与指导序列相结合，此最后一轮剪接释出一个4核苷酸产物而形成一个缩短了的新环状RNA。最后，此395-核苷酸内含子打开，生成L19 RNA。此L19 RNA是稳定的，因为它不再含有富含嘧啶的互补序列了。

然而，在L19 RNA中仍然含有一个G结合位点和一个指导序列。因而Cech认为L19 RNA有可能作用于外部的底物，并进行了大量研究。结果表明，L19 RNA以高度特异的方式催化寡聚核糖核苷酸的裂解和联结反应。五胞核苷酸(C_5)被L19 RNA转变为更长一些和更短一些的寡聚体(图12-43)。具体地说，C_5被降解为C_4和C_3。与此同时，又能生成C_6和更长一些的聚合物。可见，L19 RNA既是一个核糖核酸酶，又是一个RNA聚合酶。

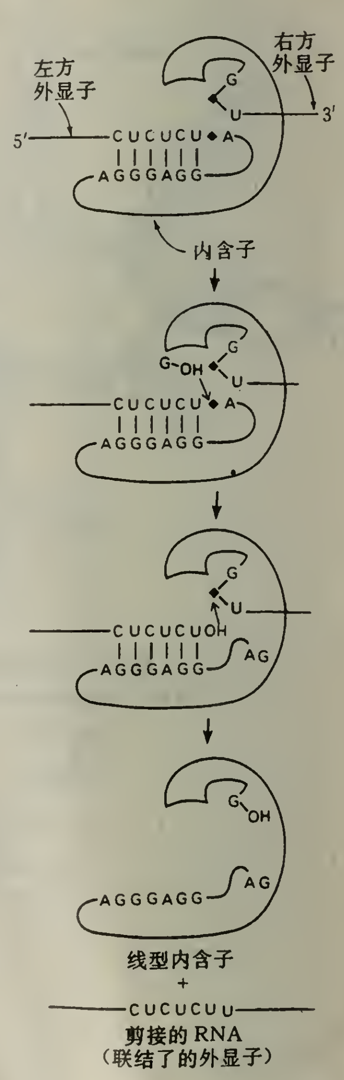


图12-42　设想的四膜虫内含子自我剪接的催化机制

引自T. Cech. RNA as an Enzyme. Copyright © 1986 by Scientific American, Inc.

此RNA性质的酶对C_5的催化作用比对U5快很多，而对A_5和G_5则完全没有作用。它按照Michaelis-Menten动力学行为，对C_5的K_m值为42 μm，kC_ot值为0.033/s。其底物必须有个2′-OH，这是由于发现脱氧-C_5是C_5的竞争性抑制剂而证明的。由此可见，L19 RNA表现出经典酶催化作用的几个标志：高度的底物特异性；Michaelis-Menten饱和动力学行为；和对竞争性抑制剂的敏感性。

此395-核苷酸RNA分子是怎样发挥核糖核酸酶和RNA聚合酶两方面的作用的？它的三维结构虽然还不清楚，但可对其催化机制做些推断(图12-43)。此底物的5个C碱基中若干个大概是与此RNA性质的酶中富含G的序列以碱基配对的方式结合起来。然后，结合在酶上的C_5的第4和第5号C之间的磷酸二酯键受到此酶的末端G残基的3′-OH的攻击。结果在酶的末端G与底物的末端C之间形成一个磷酸二酯键，从而产生一个-GpC共价中产物，与此转酯作用发生的同时C_4被释放出来了。此新生成的高度不稳定的磷酸二酯键可被水分子攻击，结果释出pC，并使酶恢复原状。以上三步反应就起到了L19 RNA的核糖核酸酶活性的作用。

另一方面，此GpC酶中产物可结合上另一个C_5分子，然后此C_5分子攻击GpC，结果产生C_6并使酶恢复原状。此步骤起到了L19 RNA的聚合酶作用。可见，共价中产物-GpC既可被OH^-攻击，亦可被第二个底物的3′-OH攻击。高pH有利于进行水解，而高浓度的C_5底物则促进转酯作用以生成C_6。连续的转酯作用可使生成的C_6延长为C_7以至更长些的寡聚体。

此 RNA 性质的酶的催化功效与蛋白质性质的酶相比情况如何？看来，L19 RNA 的催化功效没有功效最高的蛋白质性质的酶那样高，但与一般的酶相似。它催化 C_5 水解的速度约为不催化时速度的 10^{10} 倍。很明显，RNA 分子和蛋白质一样，是能够作为很有效的催化剂的。和蛋白质一样，RNA 分子是能够形成精确的三维结构以结合特异的底物和稳定暂态分子的。然而，RNA 与蛋白质的不同之处在于，它不能形成大的非极性囊袋。而且由于 RNA 是由仅 4 种结构元件(building block)构成的，蛋白质则是由 20 种构成的，所以 RNA 的多面手性能远不如蛋白质。因此，大多数的酶是蛋白质。但必须记住，蛋白质并非唯一的催化剂。现在把 RNA 性质的酶叫做核酶(ribozyme)，除了 L19 RNA 外已发现了一些其它的核酶。看来核酶更适合于识别和催化单链的核酸分子，因为核酶和这类底物用的是共同的碱基配对的语言。

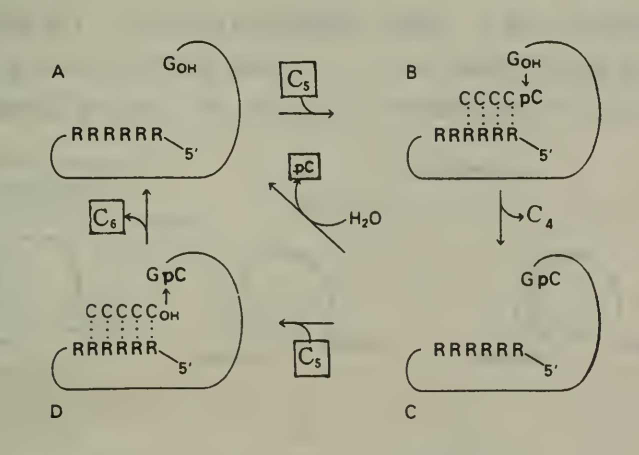


图 12-43 设想的 L19 RNA 的催化机制

A. 单独存在的酶；B. C_5 以氢键与酶连接起来，并且其末尾的 pC 与酶的末端 G 以共价键连在一起。剩余的 C_4 部分由酶上解离出来；C. GpC 水解使酶恢复为其原来状态；D. 另一条途径是，共价联结着的 pC 可被第二个 C_5 分子攻击，从而生成 C_6。引自 A. J. Zaug and T. R. Cech. Science. 1985. 231:473.

核酶的发现使得人们对于生命的起源有了新的认识。以前一般认为，由于 DNA 和蛋白质是生命的最基础物质，因而生命的最初形式必定是 DNA 或者蛋白质。然而 DNA 的复制需要蛋白质(酶)的催化，而特定蛋白质分子的合成又必须以 DNA 为模板，因而二者究竟是那一种首先出现，实在难以推断。核酶的发现使得人们普遍认为：生命的最初形式大概是 RNA，最初的生命界可能是个 RNA 王国。因为 RNA 既可作为模板而复制繁殖，同时它又是催化剂，即它兼有 DNA 和蛋白质二者的功能。然而在进化过程中，由于作为模板而遗传的功能 RNA 不如双链 DNA 稳定；而作为催化剂的功能它又不如蛋白质那样灵活多样，所以作为遗传信息携带者的功能它让位给 DNA，作为催化剂的功能则大部分由蛋白质所取代，RNA 则仅保留了它作为信使和一部分催化作用等功能，这就是目前生命界的实际情况。这种推断是否正确还有待实验的证明。在应用方面，人们正在设计合成特异切割病毒 RNA 或其它 RNA 的核酶，以便用以治疗包括艾滋病、癌症在内的疾病。这种研究虽然目前还没有应用于临床上的报道，但是是有前途的。

四、剪接体催化的剪接可能是由自我剪接进化产生的

酵母和真菌线粒体中的 mRNA 前体也发生自我剪接，单细胞生物，如单细胞藻(Chla-my-

donionas)叶绿体中的某些 RNA 前体也是自我剪接。自我剪接反应可按攻击上游剪接位点的单元的本质进行分类。I 类自我剪接是由一个鸟苷辅助因子激发的,就像前面所说四膜虫中的那种情况。Ⅱ类自我剪接的攻击单元是内含子中一个特异腺苷酸的 2′-OH(图 12-44)。转移 RNA 前体中的内含子是用一种完全不同的剪接方式剪接掉的,例如,酵母中酪氨酸 tRNA 前体中的一个 14-核苷酸内含子是先由内切酶切除,然后由连接酶将上游和下游两个片段联结在一起而剪接掉的。

I 类和Ⅱ类自我剪接在两方面与剪接体催化的剪接是类似的。第一,起始的步骤是由一个核糖的羟基攻击 5′剪接位点。然后新生成的上游外显子的 3′-OH 末端攻击 3′剪接位点,从而与下游外显子之间形成一个磷酸二酯键。第二,两步反应都是转酯作用,反应中在每个剪接位点处的磷酸基团都保留在产物中。磷酸二酯键的数目维持不变。Ⅱ类剪接还有两点与 mRNA 前体的剪接体催化的剪接方式相似。第一,头一步裂解作用是由内含子本身的一部分(A 的 2′-OH)引发的,而不是由一个外部辅助因子(G)引发的。第二,内含子是以套索的形式释放出来。

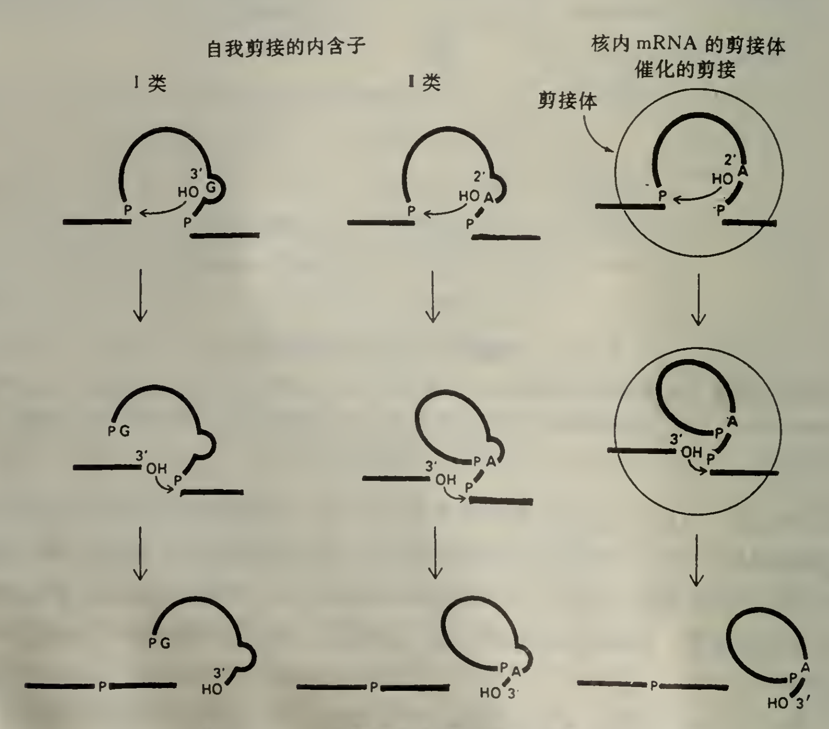


图 12-44　自我剪接和剪接体催化的剪接的比较

在 I 和Ⅱ类剪接中催化位点是由内含子自身形成的;与之相反,细胞核中 mRNA 前体的剪接是由剪接体中的 snRNAs 催化的。引自 P. A. Sharp. Science. 1987. 235:769.

Phillip Sharp 提出,mRNA 前体的剪接体催化的剪接作用是由 RNA 催化的自我剪接作用进化出来的。Ⅱ类剪接作用恰好是 I 类剪接作用和高等真核生物细胞核中的剪接作用之间的一种中间形式。由Ⅱ类剪接作用过渡到剪接体催化的剪接作用的一个主要步骤,是将催化力由内含子自身转给其它分子。剪接体的形成给予了内含子新的自由,因为它不再被迫必须提供剪接的催化中心了。用外部催化剂进行剪接的另一个优点是更易于调控剪接作用。我们的眼界已经开阔到认识更多的早期 RNA 王国与当代生命界之间的潜在连续性。

参 考 文 献

1. Lehninger, A. L., Nelson, D. L. and Cox, M. M. Principles of Biochemistry (2nd ed.). Worth Publishers, (Chapters 23.) 1993.
2. Stryer, L. Biochemistry (4th ed.). W. H. Freeman and Company, (Chapters 33 and 37.) 1995.
3. Lewin, B. Genes V, Oxford University Press, (Chapters 22, 23, 24, 25, 26, 27 and 28.) 1994.

（齐顺章）

第十三章　真核基因表达的调控

真核基因表达的调控是当前分子生物学中最活跃的研究领域之一。当内外环境发生改变时，生物常用调控某些基因表达和另些基因不表达的方式以及表达量的多少进行适应。而且细胞的生长和增殖，个体发育过程中的细胞分化，都是以基因表达的调控为基础的。本章首先介绍真核基因表达调控的一般原理，然后在此基础上介绍关于调控细胞增殖和个体发育过程的已有资料。

第一节　真核基因表达的调控

自从 60 年代 Monod 和 Jacob 用经典遗传学的方法揭示出原核基因表达的操纵子调控模式起，人们就开始企图用类似的方法研究真核基因表达的调控机制。然而直到 70 年代 DNA 重组技术出现之前没有任何实质性的进展。这主要是由于在原核生物中行之有效的遗传分析方法在真核中难以使用，加上难以获得真核生物的突变体，因而使真核生物基因表达调控机制的阐明至少比原核生物推迟了 20 年。

重组 DNA 技术的出现，使得真正的研究真核基因的结构及其表达调控的机制成为可能。例如，可以将一个特定的真核基因（基因组基因或 cDNA）分离出来，从而分析它的编码区和上游、下游侧翼（flanking）序列，然后按设计的方案进行各种改造，再送回细胞中进行表达。根据改造后表达情况的变化，可以找出调控此基因表达的序列，并进一步研究其调控机制。转基因动、植物的成功，更使我们有可能在完整生物中研究其基因表达的调控机制。由于研究真核基因表达的调控无论在理论上或实践中都有十分巨大的意义，因而近年来这方面的研究非常活跃，文章浩如烟海，已揭示出来了大量的事实，下面做简要介绍。

一、和原核相比真核基因表达调控的一些特点

真核基因表达调控的许多基本原理和原核是相同的。例如，和原核一样，真核基因表达的调控也有转录水平的调控和转录后的调控，并且也以转录水平的调控为最重要。又如，在真核结构基因的上游和下游（甚至内部）也存在着许多特异的调控成分，并依靠特异蛋白因子与这些调控成分的结合与否调控着基因是否转录。然而已经发现，在真核基因表达的调控中至少有 4 个带普遍性的与原核不同的机理，即：①原核的染色质是裸露的 DNA，而真核的染色质则是由 DNA 与组蛋白紧密结合形成为核小体（见第十二章）。在原核中染色质的结构对基因的表达没有明显的调控作用，而在真核中这种作用是明显的。②在原核基因转录的调控中，既有用激活物的调控（阳性调控或正调控），也有用阻遏物的调控（阴性调控或负调控），二者同等重要。在真核中虽然也有正调控成分和负调控成分，但迄今已知的主要是正调控。而且一个真核基因通常都有多个调控序列，必须有多个激活物同时特异地结合上去才能启动基因的转录。③原核基因的转录和翻译通常是在同一地点同时进行的（图 13-1），即在转录尚未完成之前翻译

便已开始。而真核基因的转录是在细胞核中进行，生成的初级转录物需在核中加工为成熟的mRNA，然后mRNA进入胞质，在胞质中进行翻译。所以真核基因的转录和翻译是在不同地点不同时间进行的，从而使得真核基因的表达有多种转录后的调控机制，其中许多机制是原核所没有的。④真核生物大都为多细胞生物，在个体发育过程中发生细胞分化后，不同细胞的功能不同，基因表达的情况也就不一样，某些基因仅特异地在某种细胞中表达，称为细胞特异性或组织特异性表达，因而具有调控这种特异性表达的机制。以上是现在已知的真核基因表达调控的特点，是否还有其它特点，尚待进一步发掘。

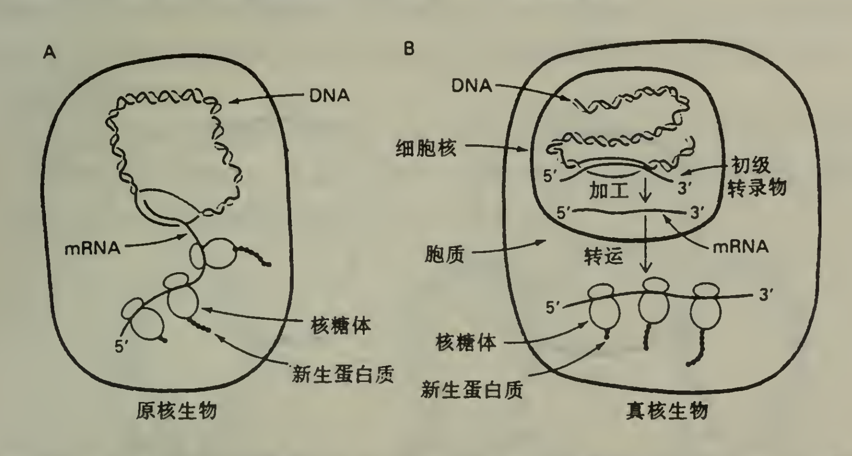


图 13-1　在原核中转录和翻译是紧密偶联的(A)；在真核中则转录和翻译是在空间和时间上都是分开的(B)

二、调控蛋白与DNA结合的方式

前已述及，基因表达调控的基本方式是靠调控蛋白与基因调控成分的特异结合，也靠调控蛋白与调控蛋白之间的特异结合，这在真核和原核生物中都是一样的。因此，在讨论真核基因表达调控的具体机制之前，先介绍一下蛋白质与DNA之间以及蛋白质彼此之间的结合方式是必要的。

调控蛋白通常是结合在特异的DNA序列上。但也与非特异DNA序列结合，不过与特异序列结合的亲和力要比非特异序列一般高10^5～10^7倍。这种区别的分子原理曾经是深入研究的课题。一个带普遍性的结论是，调控蛋白通常有独立的结合DNA的结构域。而且，这种结构域中与DNA接触的亚结构的结构基元仅有少数几种。

(一)蛋白质对特异DNA序列的识别

在讨论这些蛋白质的结构之前，先看一看能够被调控蛋白所识别的DNA表面，这是调控蛋白能与特异DNA序列相结合的分子基础。为了特异地结合，调控蛋白必须能够将它特异结合的DNA序列与其它序列区分开来，这种区分是靠暴露在DNA大沟(它易于被蛋白质接触)表面的不同碱基的基团有所不同，并且主要是靠这些基团是氢键的给体或受体的情况不同(图13-2)而识别出来。因此，在大多数蛋白质与DNA之间的接触中，其特异结合是靠氢键的形成。一个明显的例外是在嘧啶C-5附近的非极性表面，在该处由于胸腺嘧啶有个凸出的甲基而易于将它与胞嘧啶区分开来。蛋白质和DNA也能在DNA的小沟处接触，但这里可以形成氢键的花样一般不易于将不同的碱基区分开来。

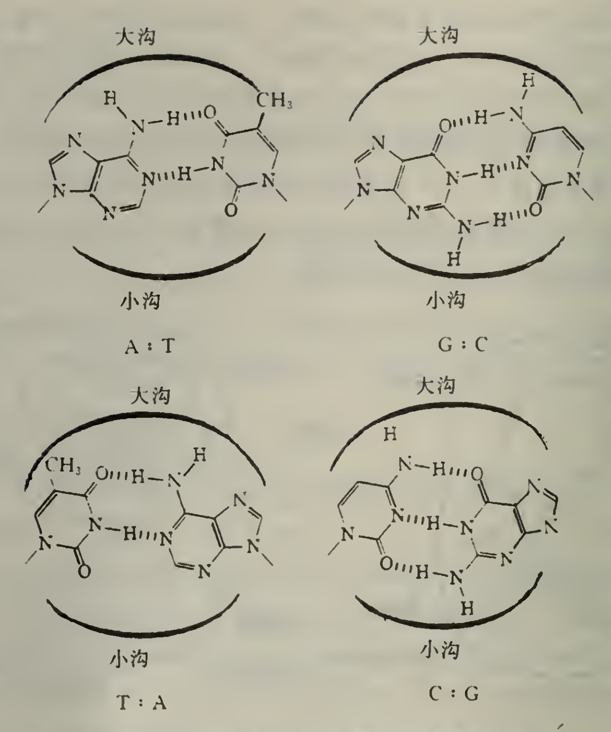


图 13-2　在 DNA 的大沟中 DNA 碱基对暴露的功能团

注意不同碱基对露出的功能团的分布位置是不同的，藉此蛋白质可将不同的碱基对识别出来。

对于调控蛋白来说，最常见的其侧链与 DNA 的碱基形成氢键的氨基酸残基有天冬酰胺、谷酰胺、谷氨酸、赖氨酸和精氨酸。有没有一个简单的"识别密码"？按此密码某个氨基酸总是与某个碱基配对？在谷酰胺或天冬酰胺与腺嘌呤的 N^6 和 N-7 位之间形成两个氢键（图 13-3 A），它们不能与任何其它碱基形成这种样子。同样，一个精氨酸残基可与鸟嘌呤的 N-7 和 O^6 形成两个氢键（图13-3B）。然而，在鉴定了许多DNA与蛋白质相结合的结构证明，一个蛋白质

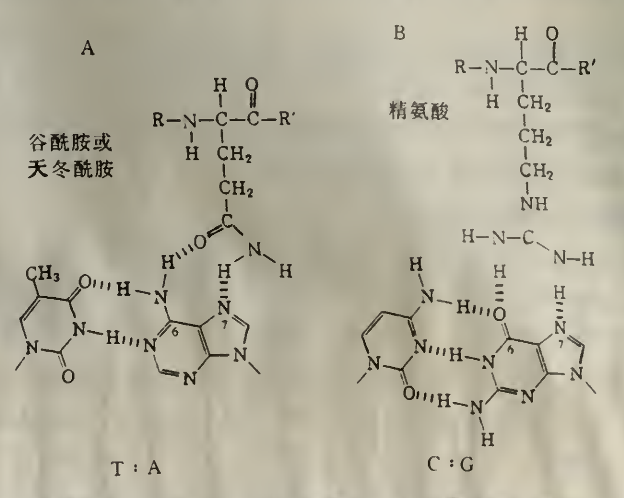


图 13-3　在 DNA 和调控蛋白相结合的结构中观察到的特异氨基酸和碱基对相互作用的两个例子

识别碱基对的方式是多种多样的，不存在简单的密码。有些蛋白质是用谷酰胺-腺嘌呤的相互作用识别 A═T 碱基对，而另些蛋白质则是用适合于胸腺嘧啶的甲基的范德华袋(van der Waals pocket)的机制识别 A═T 碱基对。现在还不可能用测定结合DNA 的蛋白质的结构来推断该蛋白质所结合的 DNA 序列。

调控蛋白结合 DNA 的结构域看来是比较小的(60～90 个氨基酸残基)。真正与 DNA 接触的仅是这些结构域中的一小套氨基酸，并且在这些氨基酸存在之处的蛋白质结构不是随意的。已发现在很多调控蛋白中有两种结构基元(structural motif)在与DNA 结合中起着主要的作用：即锌指(zinc finger)基元和螺旋-转角-螺旋(helix-turn-helix)基元。在某些蛋白质中还存在有其它的与 DNA 结合的基元，但这里着重讨论这两种研究得比较清楚的基元。

(二)锌指(zinc finger)

图 13-4 示出了锌指的结构，在其中一小组保守的氨基酸与一个锌离子结合起来，形成蛋白质中一个相对独立的结构域。有两类与 DNA 结合的蛋白质具有这种结构，即经典的锌指蛋白(zinc finger protein)和类固醇受体(steroid receptor)。

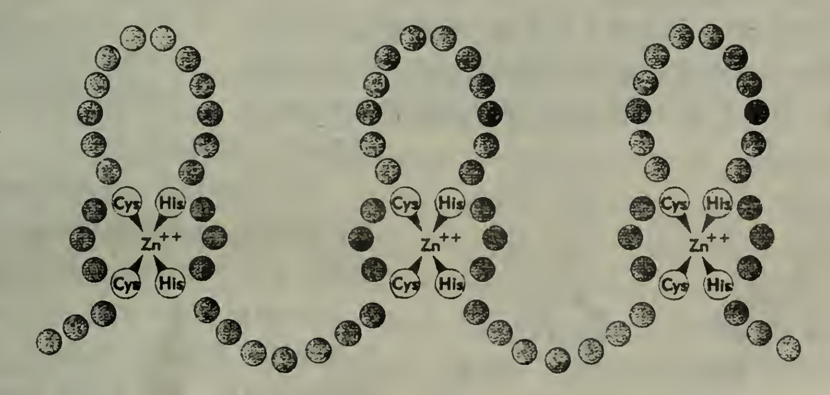


图 13-4 转录因子 SP1 有一排 3 个锌指，每个都有个特征性样式，即用半胱氨酸和组氨酸残基构成结合锌的位点

图 13-4 中所示为经典的锌指蛋白，通常有一系列锌指。单个锌指的共有序列为：

半胱-$X_{2\sim4}$-半胱-X_3-苯丙-X_5-亮-X_2-组-X_3-组，把这种基元叫做锌指是因为有个氨基酸的环，由锌结合位点凸出如手指，上述锌指又叫做半胱 2/组 2 锌指。锌被抓持在由保守的半胱氨酸和组氨酸残基形成的四面体结构中。锌指本身由约 23 个氨基酸组成，锌指间的联结物(linker)通常是 7～8 个氨基酸。

解释具有公认的锌指时必须谨慎，特别是在解释一个仅具有一个锌指的蛋白质时更是如此。因为锌指可以是与RNA 的结合有关，而不参与和 DNA 的结合，甚至与任何核酸的结合都没有关系。例如，最初鉴定的锌指蛋白，TFⅢA(RNA 聚合酶Ⅲ转录 5 S rRNA 基因时所需要的转录因子)既与 5 S 基因结合，也与转录产物 5 S rRNA 结合。翻译起始因子，eIF2β 有一个锌指，它的突变影响对起始密码子的识别。

然而，锌指是与DNA 结合的蛋白质常见的基元。锌指通常组织成为一个串联重复的样子，但也有多于一组的锌指。不同蛋白质的锌指数目不同。在 TFⅢA 中锌指的数目多至 9 个，它们几乎占了整个蛋白质的长度。在果蝇调节物 ADR1 中则仅有 2 个锌指，它们仅为蛋白质的一个小的结构域。在转录因子 SP1 的结合 DNA 的结构域中有 3 个锌指(图 13-4)。

图 13-5 为 TFⅢA 分子中 9 个锌指与 DNA 结合的模式图。锌指的尖端可进入 DNA 的大沟或小沟，以识别它特异结合的 DNA 序列并与之结合。每个锌指结合约 5 bp，相当于双螺旋的半转。9 个锌指共结合约 45 bp，与 5 S rRNA 基因的内部启动子的长度相近。5 S rRNA 基因的启动子在转录起始位点的下游，即在基因的内部。TFⅢA 特异识别此启动子序列并与之结合。在一分子 TFⅢA 与此序列结合而形成复合体后，TFⅢC，TFⅢB 和 RNA 聚合酶Ⅲ依次结合上去使 5 S rRNA 基因得以转录。

在转录时，TFⅢA 仍然结合在基因上。有多个锌指的优点之一是，当 RNA 聚合酶Ⅲ通过时，TFⅢA 的一些锌指可将 DNA 的模板链释放出来，而其余的锌指则仍结合在双螺旋上，或结合在编码链上。

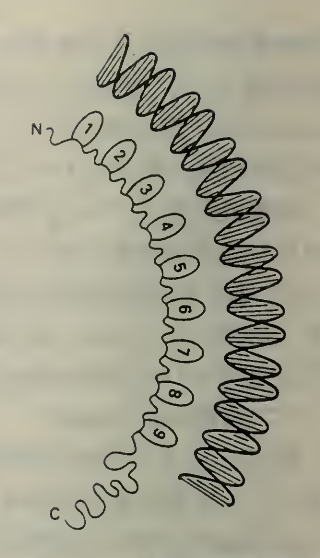


图 13-5　TFⅢA 的结构域结构
9 个锌指与 5 S rRNA 基因的内部调控区结合促进其转录

由于已知在 RNA 聚合酶Ⅱ和Ⅲ的一些转录因子中具有锌指，因而当一个蛋白质具有多个锌指时，可以认为它可能是个转录因子。用这种观点鉴定出来了在果蝇胚胎发育过程中起作用的一些转录因子。

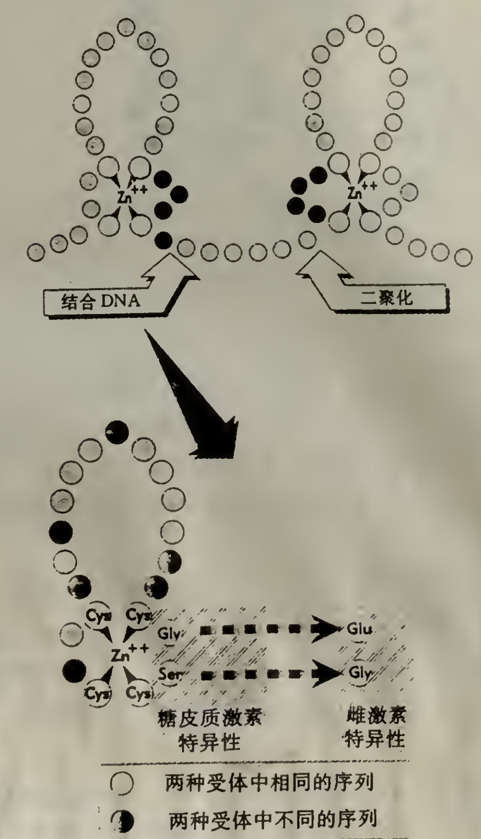


图13-6　类固醇受体的第 1 个锌指控制与 DNA 结合的特异性，第 2 个锌指控制形成二聚体的特异性

另一类锌指蛋白是类固醇受体（和一些其它蛋白质）。由于这类锌指是由两对半胱氨酸残基与 Zn^{2+}形成的，所以也叫做半胱 2/半胱 2 锌指（但也有人不把它归为锌指一类）。已知的这种锌指结合的 DNA 区比较短，并且是回文结构。糖皮质激素和雌激素受体都有两个锌指（图 13-6）。实验表明，当将类固醇受体的后两个半胱氨酸残基突变为组氨酸残基时，它就不能再活化其靶基因了。可见半胱 2/半胱 2 锌指和半胱 2/组 2 锌指是不同类型的锌指，不能交换使用它们。

类固醇激素受体是以二聚体形式发挥其促进转录作用的（详见后面）。它们的两个锌指的功能不同。第 1 个锌指的右侧是控制与 DNA 结合的，第 2 个锌指的左侧则是控制形成二聚体的能力的。第 1 个锌指控制与 DNA 结合的直接实验证据是，当将雌激素受体的此锌指缺失掉，并用糖皮质激素受体的此锌指取代时，则新蛋白质识别糖皮质激素受体的靶位点（GRE 序列），不再识别雌激素受体的靶位点（ERE 序列）了。可见，此区域规定了它识别 DNA 序列的特异性。

GRE 和 ERE 序列是很相似的，进一步的实验证明，区分这两个序列是依靠第 1 个锌指中的两个氨基酸。糖皮质激素受体和雌激素受体之间在序列上的差异主要是在锌指的基部。将按图 13-6 中指明的位置改变氨基酸时，则糖皮质激素受体不再结合 GRE，而是结合在 ERE 上了。迄今仅在真核生物中发现有锌指蛋白，在原核中尚未发现。

（三）螺旋-转角-螺旋

这种与 DNA 结合的蛋白质基元首先是在原核中发现，并且是首先详细研究过的。许多原核的调控蛋白（例如 λ 噬菌体的 Cro 蛋白，它与 DNA 的结合样式已被详细研究了）都是以这种方式与 DNA 结合的。螺旋-转角-螺旋基元的基本组成是：两个 α-螺旋被一个 β-转角隔开。此结构本身通常不稳定，但它是较大的结合 DNA 的结构域的活性部分。两个 α-螺旋之一是起识别作用的螺旋，因为它通常含有多个与 DNA 相互作用的氨基酸残基，并且被安置在大沟中。目前发现许多真核调控蛋白也含有与之非常相似的与 DNA 结合的这种基元，例如同源异型域就是如此。

同源异型域（homeodomain）最初是由调控果蝇早期发育的调控蛋白中发现的（其名称来源及一级结构等详见第三节）。编码此结构域的 DNA 序列称为同源异型框（homeobox）。在果蝇中调控其早期发育的一大类基因中的同源异型框的同源性很高，而同源异型框外的序列则很少同源性。后来发现，同源异型域广泛存在于真核生物的蛋白质中，并且不限于调控发育的调控蛋白中，在其它转录因子中也存在同源异型域，但它们的保守性不像上述果蝇中的一大类的保守性那样高。同源异型域担负着含有它的蛋白质与 DNA 结合的作用，不同的同源异型域特异的识别不同的 DNA 序列。同源异型域有 3 个 α-螺旋。图 13-7 是果蝇 *engrailed* 基因产物的同源异型域与 DNA 结合的模式图。从图中可以看出，螺旋 3 结合在 DNA 的大沟中，它是蛋白质和 DNA 的主要接触部位。螺旋 3 的一些部位与磷酸骨架结合，这是没有特异性的。但中间部分则与碱基相结合，这种结合规定着特异性。螺旋 2 和螺旋 3 形成螺旋-转角-螺旋式样，螺旋 1 和螺旋 2 为反平行样式，它们与螺旋 3 近于垂直。此外，在结构域 N 端有个伸展的臂，它伸入到小沟中，成为蛋白质与 DNA 的另一个接触位点。

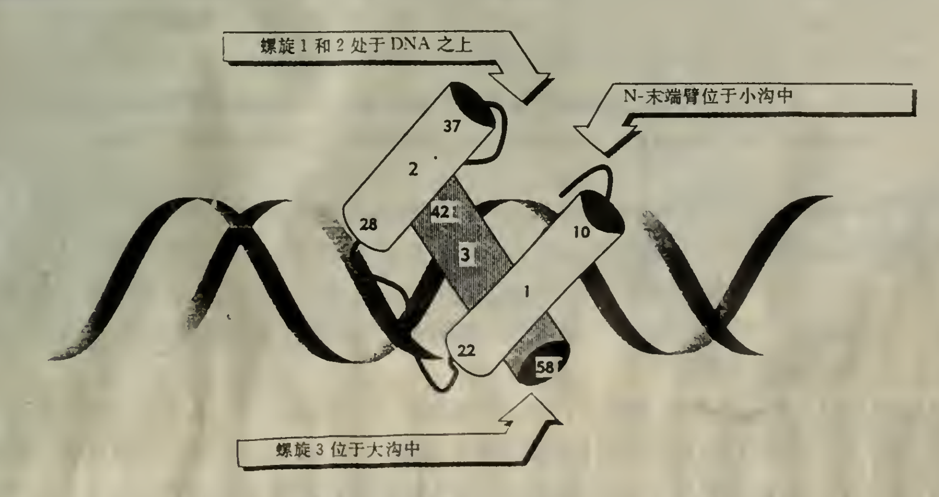


图 13-7　果蝇 *engrailed* 基因的产物与 DNA 结合的模式图

同源异型域的螺旋 3 结合在 DNA 的大沟中，螺旋 1 和 2 则居于双螺旋的外侧。螺旋 3 与磷酸骨架和特异碱基都接触。N-末端臂居于小沟中，形成额外的接触。

三、调控蛋白与蛋白质结合的方式

调控蛋白除了与 DNA 结合的结构域外，通常还有与蛋白质结合的结构域，用以和 RNA 聚合酶、其它调控蛋白相互作用，也用于相同调控蛋白分子之间的作用。调控蛋白所结合的 DNA 区常常是一小段反向重复的 DNA 序列（回文结构），因而常常是 2 个或 4 个同一调控蛋白分子结合在此序列上。例如，在原核中，Cro 蛋白是以二聚体结合在 DNA 的一段特异回文序列上，lac 阻遏物则由相同亚基组成的四聚体结合在乳糖操纵子的操纵基因上。

在真核中，许多转录因子以二聚体结合在 DNA 上。这些调控蛋白除了与 DNA 结合的结构域外，还有与蛋白质结合的结构域。这些结构域参与二聚体的形成。这些二聚体有的是由相同亚基所组成，称为同二聚体（homodimer）；有的则由不同亚基所形成，称为异二聚体（heterodimer）。二聚体的形成是它与 DNA 结合的必要条件。和蛋白质与 DNA 结合的基元一样，媒介蛋白质-蛋白质相互作用的蛋白质结构基元也可归属于几种类型。其中鉴定得比较清楚的有螺旋-环-螺旋和亮氨酸拉链两种。

（一）螺旋-环-螺旋（helix-loop-helix，HLH）

含有螺旋-环-螺旋的蛋白质称为螺旋-环-螺旋蛋白质（HLH 蛋白），它们的 HLH 的共同结构特点是：在一个 40～50 个氨基酸残基的片段中有两个双亲（amphipathic）α-螺旋，两个螺旋用长度不等的联结区（环）分开。此类蛋白质用两个螺旋的相应面上的疏水基间相互作用的方式形成同二聚体和异二聚体。螺旋区为 15～16 个氨基酸长，每个螺旋都有几个保守残基。图 13-8 比较了两个例子。形成二聚体的能力是靠双亲 α-螺旋，这对所有的 HLH 蛋白都是如此。环的作用大概只是使得两个螺旋区能够互不依赖的自由地起作用。

MyoD	Ala Asp Arg Arg Lys Ala Ala Thr Met Arg Gln Arg Arg Arg	碱性区
Id	Arg Leu Pro Ala Leu Leu Asp Gln Glu Glu Val Asn Val Leu	Id 中没有 6 个保守残基
MyoD	Leu Ser Lys Val Asn Gln Ala Phe Gln Thr Leu Lys Arg Cys Thr	螺旋 1
Id	Leu Tyr Asp Met Asn Gly Cys Tyr Ser Arg Leu Lys Gln Leu Val	MyoD 和 Id 中都有保守残基
MyoD	Lys Val Gln Ile Leu Arg Asn Ala Ile Arg Tyr Ile Gln Gly Leu Glu	螺旋 2
Id	Lys Val Gln Ile Leu Glu His Val Ile Asp Tyr Ile Arg Asp Leu Glu	

图 13-8　所有 HLH 蛋白都有螺旋 1 和螺旋 2，并用 10～24 个残基的环将它们分开

碱性 HLH 蛋白在紧邻螺旋 1 处有个具有保守正电荷的区域。下面划线者为保守残基。

从图 13-8 中可以看出，在 MyoD 的 HLH 基元附近有个高度碱性的区段，此区段是结合 DNA 所必需的，大多数 HLH 蛋白有此区段，含有此区段的 HLH 蛋白称为碱性 HLH 蛋白（basic HLH protein，bHLH 蛋白）。图中还示出，在 bHLH 蛋白中，其 14 个氨基酸的碱性区段内有约 6 个残基是保守的。而有些 HLH 蛋白（如图中的 Id）则缺乏此碱性区段，或者含有破坏此碱性区段功能的脯氨酸残基。

已知的 bHLH 蛋白有：E12 和 E47，它们是结合在一个免疫球蛋白基因的增强子中的一个元件（element）上的两种蛋白质；MyoD、肌细胞生成素（myogenin）和 Myf-5 都是参与肌细胞生成（myogenesis）的转录因子；*myc* 基因类[它们是致癌基因的细胞对应物（counter-part）]的产物，它们参与生长调节；和果蝇中一类规定神经系统发育的因子。bHLH 蛋白按其表达情况可分为两类：一类是在各种组织中广泛表达的蛋白质，其中有哺乳动物的 E12，E17 和果蝇的

da 蛋白；另一类是组织特异性表达的，如哺乳动物的 MyoD 和果蝇的 AC-S。组织特异性的 bHLH 蛋白类可能是转录调节物，它们和广泛表达类的蛋白质形成异二聚体时最有效。

bHLH 蛋白是以二聚体与 DNA 结合并发挥作用的。一个同二聚体或一个异二聚体的两个亚基都有碱性区时才能与 DNA 结合。HLH 结构域的作用大概是正确安置这两个（由每个亚基提供一个）碱性区的位置。但由不同 bHLH 蛋白形成的二聚体与 DNA 结合的能力有所不同。例如，E47 的同二聚体，E12-E47 的异二聚体和 MyoD-E47 的异二聚体都能有效地形成，并有较强的与 DNA 结合的能力；E12 的同二聚体也能有效地形成，但与 DNA 的结合很差；而 MyoD 的同二聚体本身就不易形成。为什么会有这种差异，目前还不清楚，即我们还不能根据 HLH 的序列来推断二聚体是否易于形成以及它们与 DNA 结合的强度。

与 DNA 结合的能力不同是因为 HLH 基元内部或其邻近区域的性质不同之故。例如，和 E47 不同，E12 的碱性区附近有抑制区，它抑制 E12 的同聚物与 DNA 结合。有些 HLH 蛋白，例如，前面提到的 Id，没有碱性区和/或含有破坏其功能的脯氨酸残基。此类蛋白质形成二聚体的能力与 bHLH 蛋白相同，然而在一个二聚体中含有一个这样的亚基，它就不再能够与 DNA 特异地结合了。这有力地证明了，在与 DNA 结合的蛋白质中，其与 DNA 结合的基元加倍是非常重要的。而且已知，在不同条件下形成不同的二聚体（能够和不能与 DNA 结合的）是调控基因表达的重要方式之一，因为调节物蛋白调控基因表达的先决条件是它必须与 DNA 特异结合。

（二）亮氨酸拉链（leucine zipper）

此种结构基元的特点是蛋白质的 α-螺旋的一侧集中了许多疏水氨基酸，两分子蛋白质的这种疏水侧面相互作用使之形成二聚体。这些 α-螺旋的一个突出特点是频繁出现亮氨酸，并且趋于每 7 个氨基酸残基出现一个亮氨酸，这种出现频率使得在形成 α-螺旋时，亮氨酸出现在 α-螺旋的疏水一侧，并且直线排列，如图 13-9 所示。早期设想的这些 α-螺旋之间的蛋白质-蛋白质相互作用的模式是亮氨酸残基的交错插入，因而把它叫成亮氨酸拉链。现在已知，在两个蛋白质上的亮氨酸残基是肩并肩地排列起来的，并且相互作用的两个 α-螺旋是彼此缠绕在一起，成为螺旋了的螺旋（coiled coil）。

亮氨酸拉链蛋白在序列上的另一个共同特点是，在拉链区的氨基端有个约 30 个残基的碱性区（富含赖氨酸和精氨酸）。此区的作用是与 DNA 结合，它也形成 α-螺旋。亮氨酸拉链区的作用是将一对与 DNA 结合的区域拉在一起，以结合两个相邻的 DNA 序列。两个亮氨酸拉链蛋白形成二聚体时构成 Y 字形，螺旋了的螺旋是 Y 字的干，而碱性区则为其臂。每个碱性区的 α-螺旋看来是发生弯曲，以便缠绕在 DNA 的大沟中。亮氨酸拉链蛋白在真核中广泛存在，在原核中也已有所发现。

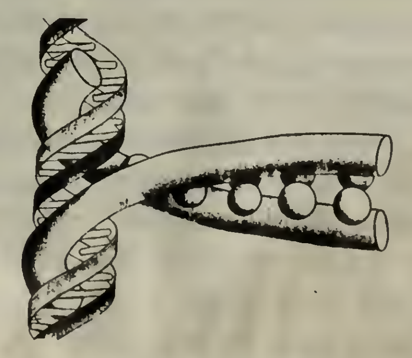


图 13-9　表明亮氨酸拉链蛋白结合在一个回文的靶 DNA 序列上的模式图

引自 C. R. Vinson，P. B. Sigler and S. L. Mcknight. Science. 1989. 246：911. ○代表亮氨酸。

酵母中的 GCN4 是个亮氨酸拉链蛋白，它是酵母细胞在对氨基酸饥饿发生应答时合成的。GCN4 同步活化 40 多个基因的转录，这些基因都编码氨基酸合成的酶。GCN4 二聚体激活物结合在 9 bp 的一个回文序

列上，此序列位于氨基酸合成的基因的上游。GCN4 有个酸性的转录活化结构域，此结构域是与碱性的结合 DNA 区以及亮氨酸拉链区分隔开的（图 13-10）。当结合在 DNA 上时，此活化结构域与 TATA 框处的转录因子ⅡD(TFⅡD)相互作用而启动转录（详见后）。

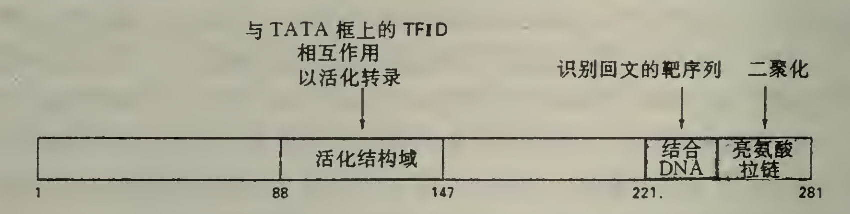


图 13-10　酵母中一个转录激活物，GCN4 的结构域结构

GCN4 促进编码氨基酸合成的酶类的基因表达。

在高等真核生物中，亮氨酸拉链蛋白媒介环状 AMP(cAMP)对转录的效应。cAMP 调控的基因含有一个 cAMP 应答元件(CRE)，它是个回文的 8 bp DNA 序列，其共有序列为：

5′-TGACGTCA-3′

3′-ACTGCAGT-5′

一个叫做cAMP 应答元件结合蛋白(CREB)的 43 kD 蛋白质结合在 DNA 的此靶序列上。两个 CREB 分子以亮氨酸拉链的方式形成二聚体，从而把两个分子中结合 DNA 的碱性区域拉在一起，并将它们定位成与回文的 CRE 结合。cAMP 是怎样调控 CREB 的转录活性的？cAMP 是由 G 蛋白级联产生的，它活化蛋白激酶 A(PKA)。PKA 是个多功能激酶，它催化 CREB 磷酸化，CREB 磷酸化后增强了二聚化作用，从而提高了它作为一个转录激活物的效应。

亮氨酸拉链蛋白可形成同二聚体或异二聚体，GCN4 和 CREB 都是同二聚体的例子。哺乳动物的转录因子 AP1(激活物蛋白 1)是一组异二聚体。AP1 的组成成分有 2：其一是基因 *c-jun* 的表达产物 Jun，*c-jun* 是动物细胞中与病毒的癌基因 *v-jun* 的对应物。后来在动物细胞中还发现有基因 *junB*，*junD*，它们与 *c-jun* 的序列同源性很高。所以现在已知有 3 种序列相似的 Jun 蛋白，它们都有亮氨酸拉链，能形成同二聚体，也能形成异二聚体。AP1 的另一组成成分是 Fos，它是 *c-fos* 基因的产物，*c-fos* 也是病毒癌基因 *v-fos* 的细胞对应物。在细胞中也有一组与 Fos 类似的蛋白质，除 Fos 外，其它蛋白质都叫做 Fos 相关抗原(Fos-related antigens，FRA)，它们组成一个 Fos 样蛋白家族。

Fos 也有亮氨酸拉链，Fos 不能形成同二聚体，但 Fos 能与 Jun 形成异二聚体。事实是，Jun-Fos 异二聚体比 Jun-Jun 同二聚体更稳定。Jun-Fos 和 Jun-Jun 识别的是相同的 DNA 上的靶位点，即 AP1 结合位点。

看来，多种亮氨酸拉链蛋白可用各种不同的组合形成很多种不同的同二聚体或异二聚体。例如，ATF-CREB 家族有 13 个成员，它们间可以彼此二聚化形成 13 种同二聚体和 78 种异二聚体。亮氨酸拉链家族生动地表明，用组合的结合(combinational association)可产生不同的调控蛋白，它们能阅读很多种不同的 DNA 序列。螺旋-环-螺旋蛋白的情况也是如此。

四、活泼转录区染色质的结构发生改变

前已述及，和原核不同，真核染色体是有高度组织结构的。这种结构对基因的表达有没有调控作用？回答是肯定的。因为现在虽然还不清楚这种结构调控基因表达的机制是怎样的，但

已有许多实验表明活泼转录区的核小体结构与非转录区的结构有所不同。而且用体外再造的染色质所做的实验结果表明，这种结构对基因的表达有着明显的影响。

说明染色质中转录活泼区结构发生改变的一个明显事实是，这些区对核酸酶的敏感性提高了。当用核酸酶，例如 DNase I 消化未发生转录的染色质时，得到的是约 200 bp 长的片段，这是由于核酸酶仅能作用于核小体核心部分以外的 DNA 特异位点上，而核心部分由于受到组蛋白的保护作用而不被消化，此结果反映出染色质的重复的核小体结构。然而在活泼转录区形成的片段更小，而且长短也不均一，说明这些区有对核酸酶高度敏感的序列，称之为超敏感位点。仅在发生基因表达的细胞中才能发现有这种对核酸酶超敏感的位点，可见它与基因的转录有关。已发现，这些位点通常位于被转录基因的 5′端 1 000 bp 的侧翼内，但也有位于离 5′端更远一些、3′端附近和甚至基因内部的。许多超敏感位点相当于已知的调控蛋白所结合的位点。现在看来，这些位点之所以对核酸酶敏感是由于该处的核小体结构已不完整，使核酸酶易于起作用之故。这种核小体的变为不完整可能是使调控蛋白易于结合，或者是由于调控蛋白的结合引起了核小体结构的不完整。

正在转录的染色质处 DNA 的另一个变化是其甲基化程度降低了。已知在真核 DNA 中的 CpG 序列的胞嘧啶是普遍甲基化的。然而许多基因在表达它们的组织中，其附近的 CpG 序列的甲基化程度低于不表达它们的组织中。

在转录时观察到的另一个明显变化是组蛋白和有关蛋白质的改变。活泼转录的染色质趋于缺乏组蛋白 H1，而且其它的核心组蛋白则被乙酰基化或与泛素(ubiquitin，一种蛋白质，它的已知功能是它与某种蛋白质结合后，使该蛋白质易于降解，即给需要降解的蛋白质打上标签)相结合而被修饰。有些情况下，在非常活泼的转录区，例如，许多真核细胞的 rRNA 基因处是没有核小体结构的。所有这些事实提示，活泼转录的染色质是通过去掉潜在的结构障碍而为转录做好准备的，但发生这些变化的机制却不甚明了。

在体外将组蛋白加到真核 DNA 中可实现染色质的再造。例如，将等分子的 H2A，H2B，H3 和 H4 加到 DNA 中可形成核小体。最近用这种再造体系进行的体外转录实验的结果可以看出，染色质结构对基因的表达有明显的影响。再造的核小体比裸露 DNA 转录的速度慢得多，如果再加上 H1(完全再造的染色质)，则转录的速度更慢。然而当将转录因子加到完全再造的染色质中时，可使转录的速度提高 100～200 倍，达到体内的水平，但对裸露 DNA 的转录速度仅能提高约 10 倍。这些结果提示，组蛋白对基因的转录在一定情况下起阻遏作用；在另种情况下又起激活作用，转录因子可能是引起基因的启动子区核小体结构发生明显改变。这种实验的另一个有意义的发现是关于增强子的效应。在体内转录时增强子的效应不受它与被转录基因的距离的影响，但用裸露的 DNA 在体外转录时则不成，只能是用完全再造的染色质时才能实现体内观察到的增强子的效应。

五、大多数真核启动子是受正调控的原因

一般说来，真核 RNA 聚合酶对其启动子只有很小或者没有特异的亲和性，转录的启动几乎总是依靠一个，而更经常的是多个激活物蛋白的作用。尽管真核的调控蛋白在某些条件下可以是激活物或阻遏物，但广泛应用的是正调控，负调控成分看来不普遍。

真核基因的表达普遍采用正调控的原因可能有二：①由于真核基因组比原核的大很多，因而调控蛋白在 DNA 上的非特异性结合便成为一个很重要的问题。随着基因组的加大，一个特

异结合蛋白质的DNA序列在不适当位置随机出现的可能性也加大。在这种情况下，能够保证调控基因特异表达的方法之一就是使用多个调控蛋白，因为几个分别特异结合几种不同调控蛋白的DNA序列，能够以适当的有功能的方式排列在一起的随机发生的可能性几乎是没有的。然而用多种蛋白质调控一个基因的表达必须都是正调控才能做到特异的调控，因为如果是用几种阻遏物进行调控，则只要结合上一种即足以封阻RNA聚合酶的作用。但是由于是用几种正调控蛋白质，则这些蛋白质必须同时分别结合在DNA特异序列上，并且然后形成复合体才能启动转录，这样就保证了基因的特异表达。现在看来，一个多细胞生物基因的调节位点（即调控蛋白特异结合的序列）平均至少是5个；②真核基因普遍采用正调控的第二个可能的原因是，由于真核基因所含基因的数目很多，但在分化了的不同细胞中，每种细胞只需表达一小套基因。例如，人的基因组中大约含有100 000个基因，但在不同的每种细胞中却只分别表达其中不同的一小部分。如果是用负调控，则任何细胞都需要合成100 000种不同的阻遏物蛋白，而且它们的浓度都必须达到足以特异地结合在调控位点上，以封阻每个基因的表达。而采用正调控，则由于每种细胞的大部分基因通常是不表达的（即RNA聚合酶不结合在其启动子上），表达的仅为一小部分，因此，每种细胞仅合成它需要的一小套激活物蛋白就可以了，这就大大减轻了细胞合成蛋白质的负担。除上述两种可能的原因外，真核基因普遍采用正调控是否还有其它的原因，尚需进一步探讨。

六、真核细胞中有3种RNA聚合酶

在原核中，RNA是由一种聚合酶合成的。而在真核的细胞核中则有3种RNA聚合酶，它们的模板特异性、所处位置和对抑制剂的敏感性是彼此不同的（表13-1）。RNA聚合酶Ⅰ位于核仁中，它在核仁处转录前后排列的一批18 S，5.8 S和28 S核糖体RNA的基因。另一个核糖体RNA分子（5 S rRNA）和所有的转移RNA分子是由RNA聚合酶Ⅲ合成的，此酶位于核浆中。信使RNA的前体是由RNA聚合酶Ⅱ合成的，此酶也位于核浆中。一些小的RNA分子，例如，剪接装置中的U1 snRNA也是由聚合酶Ⅱ合成的。

表13-1　真核的RNA聚合酶

类型	位置	转录产物	α-鹅膏蕈碱的效应
Ⅰ	核仁	18 S，5.8 S和28 S rRNA	不敏感
Ⅱ	核浆	mRNA前体和snRNA	强烈抑制
Ⅲ	核浆	tRNA和5 S rRNA	高浓度抑制

和原核的一样，真核RNA聚合酶在催化RNA合成时也是以核糖核苷三磷酸为底物，按DNA模板链的指令，由5′→3′方向一个核苷酸一个核苷酸地联结上去。真核RNA聚合酶也不需要引物，它也没有核酸酶活性，这意味着初始RNA中的错误不被校正。

三种真核RNA聚合酶都是大的蛋白质，分子量为500 kD或更大，由8～14个亚基组成。提纯的酶可按模板转录出RNA，但不能选择性地从启动子处开始。关于亚基的鉴定对RNA聚合酶Ⅱ做的比较多，其中唯一把所有亚基都确定了的是酵母的RNA聚合酶Ⅱ。真核RNA聚合酶Ⅱ有3个特性是保守的：①它们含有8～12个亚基，酵母的由11个亚基组成；②它们含有两个大的亚基（分别约为220 kD和140 kD），命名为RPB1和RPB2，它们相当于原核核心酶中起催化作用的β亚基和β′亚基。其最大（220 kD）亚基的羧基末端结构域（carboxy1-termi-

nal domain,CTD)是值得注意的,它由共有序列:酪-丝-脯-苏-丝-脯-丝的多个重复所组成,此序列是RNA聚合酶Ⅱ独有的。在酵母中有27个这种CTD序列,在哺乳动物中有50多个。CTD序列是由磷酸化作用调节的,它们参与识别活化信号。此区域是非常重要的,因为发现含有突变的RNA聚合酶Ⅱ的酵母中,当此酶的CTD序列少于10个时是不能存活的;③所有的RNA聚合酶Ⅱ都含有3个共同的亚基(RPB5,6和8)。而且这些亚基在RNA聚合酶Ⅰ和Ⅲ中也存在。线粒体和叶绿体的RNA聚合酶与细菌的相似,而不与任何核中的酶相似。

世界上每年都有上百的人死于食用毒蘑菇,尤其是*Amanita phalloides*。此种毒蘑的毒素之一叫做α-鹅膏蕈碱(α-amanitin),它是个环状八肽,对RNA聚合酶Ⅱ有强烈的抑制作用,从而封阻了mRNA前体的合成。较高浓度(1 μmol/L)的α-鹅膏蕈碱抑制RNA聚合酶Ⅲ,但聚合酶Ⅰ对此毒素不敏感。

七、真核启动子和转录因子

和细菌的RNA聚合酶一样,在启动转录时真核RNA聚合酶也需要辅助因子,并且在随后的转录过程中就不需要了。然而辅助因子的作用则真核的与原核的有所不同。在细菌中是RNA聚合酶自己识别启动子成分,但在真核中主要是由辅助因子识别启动子的组成序列,而不是RNA聚合酶本身。

真核的3种RNA聚合酶的启动子是不同的,但它们的共同特点是都含有特定的一套短的保守序列,每个序列都由特定的蛋白因子所识别并与之结合,然后这些蛋白因子与RNA聚合酶相互作用以形成起始复合体,由此复合体启动转录。RNA聚合酶Ⅰ和Ⅲ的启动子的这种保守序列的数目较少,所需要的辅助因子也比较少。RNA聚合酶Ⅱ的启动子的这种序列很多,变化也很大,所需的辅助因子当然也就很多。由此可见,真核启动子的定义和原核启动子的定义是有所不同的。细菌的启动子是由RNA聚合酶本身所识别并结合的,所以可定义为RNA聚合酶结合区,此区域在转录起始位点附近。启动子的附近虽然也有对启动子起调控作用的序列(如操纵基因),但一般不把它们看成是启动子的一部分。而真核启动子的多个保守序列都不是由RNA聚合酶本身所识别和结合的,因而不能定义为RNA聚合酶结合区。从实用出发,真核启动子可定义为包括所有那些位于转录起始位点附近的序列,这些序列都是启动转录所必需的。而把识别并结合在这些序列上的蛋白因子都称为转录因子。

需要说明的是,上述关于真核启动子的定义中没有把增强子序列包括在内,因为增强子序列可以距离转录起始位点很远。然而增强子是参与启动转录的序列,因此上述定义并非严格的定义,我们也不一定要把启动子和增强子绝对地分开。增强子并非RNA聚合酶Ⅱ所独有,RNA聚合酶Ⅰ的启动子也受增强子的协同作用,当然它们是与RNA聚合酶Ⅱ的增强子不同的增强子。此外,转录因子也不一定都直接结合在DNA上。有些转录因子结合在其它转录因子上以形成转录装置。因此,一个转录因子的决定性指标是看它的功能,即发生转录所必需的蛋白质才是转录因子。

(一)RNA聚合酶Ⅰ的启动子

在真核细胞的核中RNA聚合酶Ⅰ的启动子样式最少,RNA聚合酶Ⅰ仅由一种样子的启动子转录rRNA的基因。图13-11表示出此启动子的结构以及与启动转录有关的过程。对人类细胞中的RNA聚合酶Ⅰ的启动子研究得最清楚,它由起始位点上游区域中的两个序列组成。核心启动子(core promoter)在起始位点周围(−45～+20),它足以启动转录,然而它的效

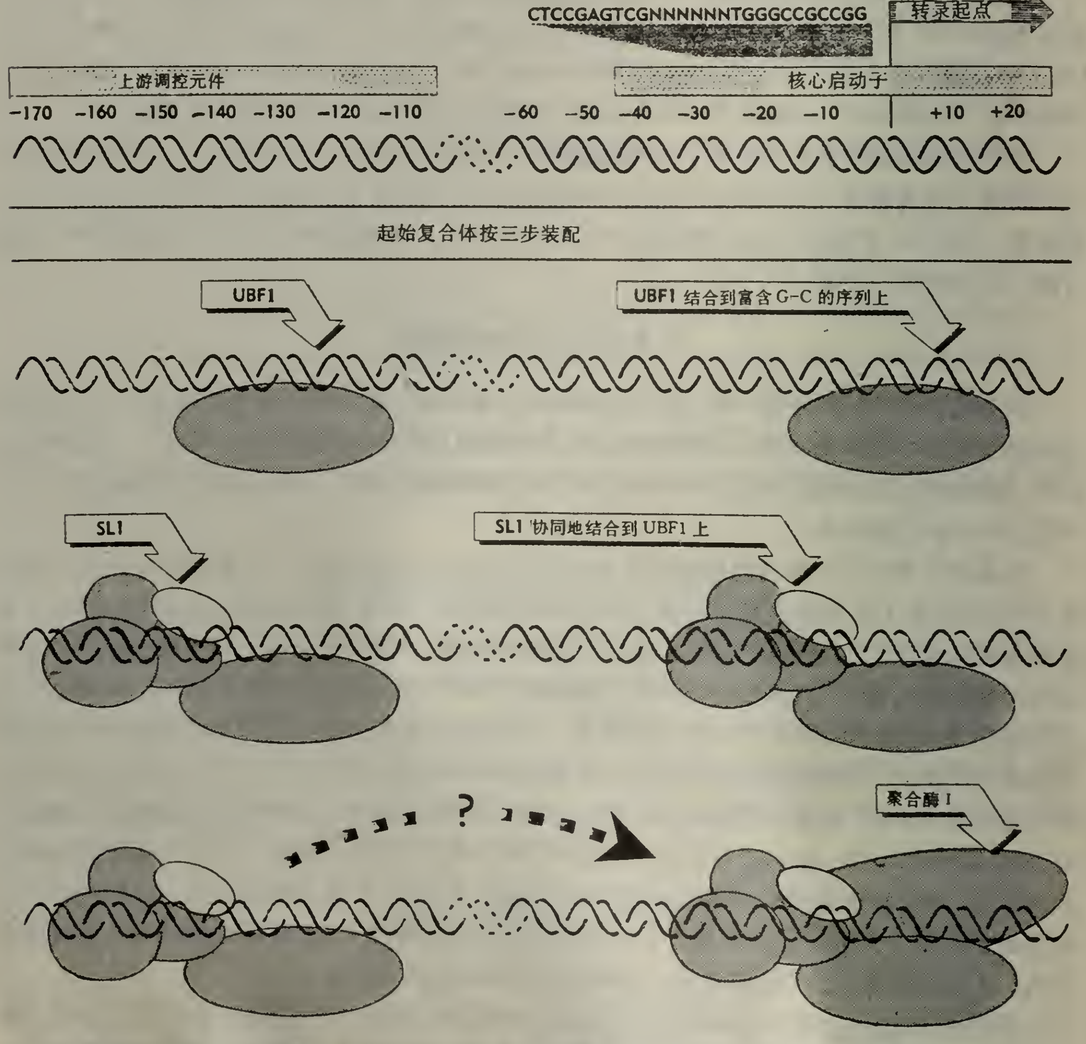


图 13-11　RNA 聚合酶 I 的启动子由一个核心启动子组成，它与上游调控元件用 70 bp 分开。UBF1 与两个区都结合，随后 SL1 能够结合。再后 RNA 聚合酶 I 与核心启动子结合

率被上游调控元件(upstream control element,UCE)大大地提高。UCE 位于－180～－107。两个区都各有一个富含 GC 的特异序列，二者约 85％相同。

RNA 聚合酶 I 需要两个辅助因子：UBF1 和 SL1。UBF1 能结合在核心启动子和 UCE 上。SL1 自己不能结合在启动子上，然而一旦 UBF1 结合上了，SL1 便可与之协同地结合在扩大了的 DNA 区域上。一旦两个因子都结合了，RNA 聚合酶 I 便能结合到核心启动子上以启动转录。可以设想结合在核心启动子上的辅助因子是与 RNA 聚合酶 I 直接相互作用的，但还不清楚结合在 UCE 上的相同因子是怎样促进转录的。后面我们将要谈到关于 RNA 聚合酶 Ⅱ 启动子的类似情况。

UBF1 是个单链多肽，它特异地结合在核心启动子和 UCE 的富含 GC 的元件上。UBF1 和

RNA 聚合酶 I 能对异源的模板有效地作用，例如，小鼠的 UBF1 因子和聚合酶 I 能识别人的基因。SL1 因子具有种属特异性，小鼠 SL1 不能在人的模板上启动转录，反之亦然。

SL1 由 4 个蛋白质组成。其中之一叫做 TBP（名称来源见后），它也是 RNA 聚合酶 Ⅱ 和 Ⅲ 启动转录所需的一个因子。RNA 聚合酶 I 所需的 TBP 在品种间是很保守的，它不负责种属特异性，也不与富含 GC 的 DNA 序列特异结合。可见种属特异性以及和 DNA 结合都是 SL1 其它组分的任务。看来 TBP 是与 RNA 聚合酶相互作用的。SL1 的行为类似于细菌的 σ 因子。分离出来的 SL1 不特异地结合到启动子上，但与其它蛋白质结合时特异地结合于启动子区。在保证将 RNA 聚合酶定位于适当的起始位点处它大概负有主要的责任。下面将要谈到，由 TBP 和其它蛋白质结合而成的因子对 RNA 聚合酶 Ⅱ 和 Ⅲ 具有类似的功能。可见所有 3 种 RNA 聚合酶在启动转录中的一个共同特点是依靠一个定位因子（positioning factor），此因子是由 TBP 与其它蛋白质结合而成的，而其它蛋白质则是对每种启动子特异的。

（二）RNA 聚合酶 Ⅲ 的启动子

RNA 聚合酶 Ⅲ 的启动子有两种。5 S rRNA 和 tRNA 基因的启动子在起始位点的下游，即在基因的内部，称为内部启动子（internal promoter）。这是很特殊的，因为其它 RNA 聚合酶以及 RNA 聚合酶 Ⅲ 的另一种启动子都是在起始位点的上游。RNA 聚合酶 Ⅲ 的另一种启动子是 snRNA（核内小 RNA）基因的启动子。图 13-12 示出了 RNA 聚合酶 Ⅲ 的 3 种类型的启动子，其中上面两个（1 型和 2 型）属于内部启动子，下面的 1 个（3 型）为上游启动子。两种内部启动子都是由两个短序列元件（sequence element）构成的，两个元件用可变的序列分隔开。1 型由框（box）A 和框 C 所组成，2 型由框 A 和框 B 所组成。2 型中框 A 和框 B 的距离可变化很大，但不能相距很近，否则将丧失功能。上游启动子的结构将在后面叙述。

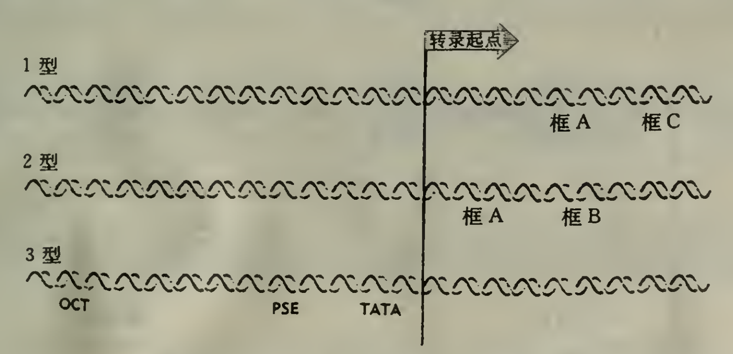


图 13-12　RNA 聚合酶 Ⅲ 的启动子类型

图 13-13 概括了内部启动子在启动转录时所发生的过程。在此过程中有 3 个辅助因子参与，即 TF Ⅲ A，TF Ⅲ B 和 TF Ⅲ C。在 1 型启动子首先是 TF Ⅲ A 结合在包含框 C 的一个序列上，这样可使 TF Ⅲ C 结合在其下游。TF Ⅲ C 的结合能使 TF Ⅲ B 结合在起始位点的周围，TF Ⅲ B的结合是伴随另一个 TF Ⅲ C 的结合发生的，最后 RNA 聚合酶 Ⅲ 结合在起始位点周围以启动转录。在 2 型启动子则是 TF Ⅲ C 直接结合在框 B 和框 A 处，然后 TF Ⅲ B 就结合上去了。

由于体外实验表明，当 TF Ⅲ A，TF Ⅲ B 和 TF Ⅲ C 结合在启动子上后，将 TF Ⅲ A 和TF Ⅲ C 移去时，TF Ⅲ B 仍能结合在起始位点附近，而且它的存在就足以使 RNA 聚合酶 Ⅲ 结合在起始位点处，可见 TF Ⅲ B 是 RNA 聚合酶 Ⅲ 所需要的唯一真正启动因子。然而 TF Ⅲ B 自己不能结合在正确的 DNA 位点上，要靠 TF Ⅲ A 和 TF Ⅲ C 的帮助，所以 TF Ⅲ A 和 TF Ⅲ C 是组装因子

(assembly factor)，TF Ⅲ B 是使 RNA 聚合酶正确定位的定位因子。像 RNA 聚合酶 Ⅰ 在启动转录时所需的 SL1 一样，TF Ⅲ B 中也含有蛋白质 TBP，TBP 是直接与 RNA 聚合酶起作用的亚基。

RNA 聚合酶 Ⅲ 的第 3 型启动子是在起始位点的上游，在图 13-13 中所举的例子是由 3 个上游元件组成的，即 TATA，PSE 和 OCT。有些 snRNA 基因是由 RNA 聚合酶 Ⅱ 转录的，另一些则由 RNA 聚合酶 Ⅲ 转录。RNA 聚合酶 Ⅱ 转录的 snRNA 基因的启动子元件也有 RNA 聚合酶 Ⅲ 的这些元件。RNA 聚合酶 Ⅲ 的上游启动子仅有 TATA 元件时也能启动转录，然而 PSE 和 OCT 元件的存在大大提高转录的效率，结合在这些元件上的因子是协同的相互作用的。

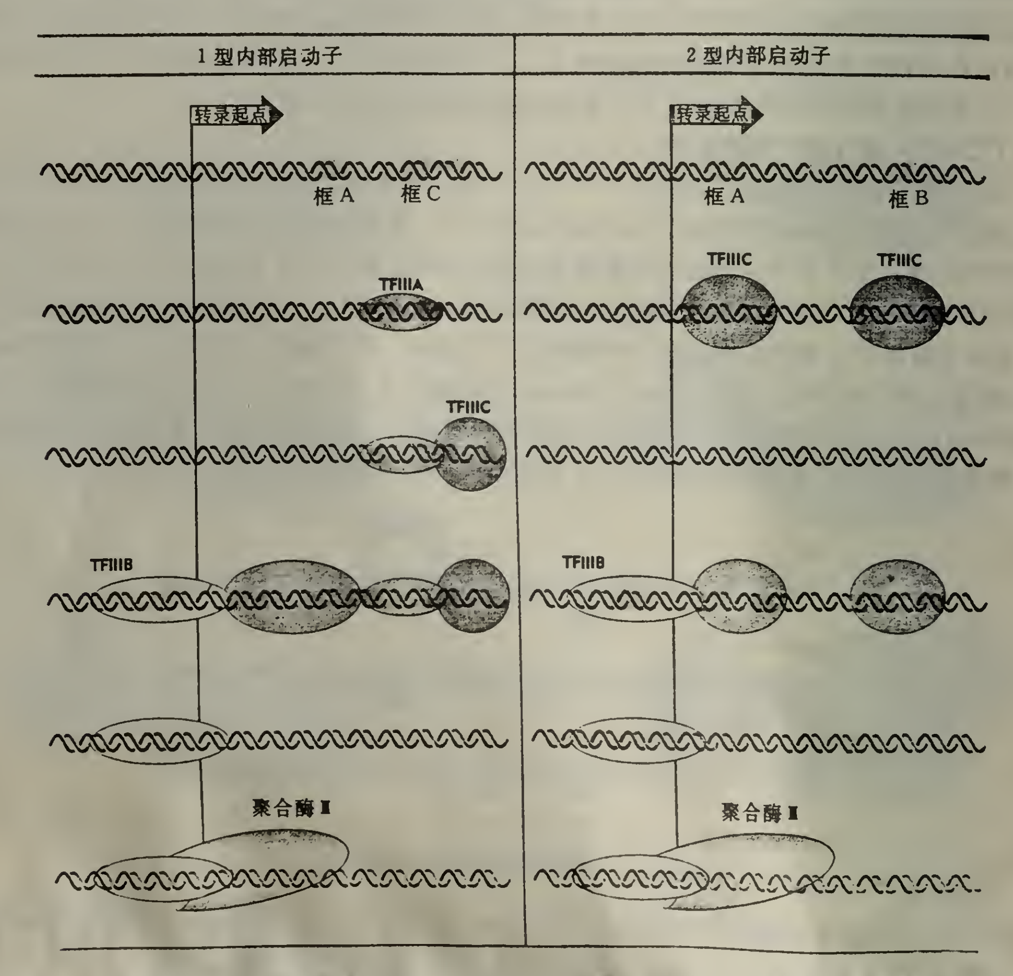


图 13-13　RNA 聚合酶 Ⅲ 内部启动子的启动过程需要装配因子 TF Ⅲ A 和 TF Ⅲ C，启动因子 TF Ⅲ B 和 RNA 聚合酶 Ⅲ

注意指明每个因子所结合的位置是它们保护的 DNA 区，而不是用分子的作用确定的。例如，图中所指出的是 TF Ⅲ C 的保护区，但可能是单个或多个蛋白质所起的作用。

结合在关键的 TATA 元件上的是个含有 TBP[TBP 是 TATA-binding protein(TATA-结合蛋白)的简称]的因子，此因子中的 TBP 与前面提到的 TF Ⅱ B 和 SL1 因子中的 TBP 是相似的，但该因子中的其它蛋白质则是特有的。TF Ⅲ B 和 SL1 都不是自己直接结合在特异的

DNA 位点上，而是由其它因子协助结合上去的。但在有 TATA 元件的启动子中，则是由 TBP 直接识别并特异地结合到 DNA 上的。现在看来，作用于 RNA 聚合酶Ⅲ的启动子的因子有个基本一致的功能，即 RNA 聚合酶Ⅲ自己不能特异地结合在 DNA 的一定位点上，而是先由这些因子结合在特定的 DNA 位点上，形成一个前起始复合体(preinitiation complex)，然后 RNA 聚合酶Ⅲ识别此复合体并结合上去成为起始复合体，在其中 TBP 是与 RNA 聚合酶直接起作用的蛋白质。

(三)RNA 聚合酶Ⅱ的启动子和转录因子

RNA 聚合酶Ⅱ所使用的启动子结构很复杂，不同启动子在序列上的变化很大。然而它们都是由多个短的序列(称为元件，element)所构成。这些元件都在起始位点的上游，通常是分散在一个大于 100 bp 的区域内。这些元件与元件之间的序列本身是不重要的，但用它们将元件与元件分隔开则可能是重要的。不同启动子所用的元件的种类、数目以及这些元件所处的位置都不同，尽管有些元件可出现在不同的启动子中。RNA 聚合酶Ⅱ自己不能识别这些元件并与之结合(所以不能按原核的定义把真核启动子也定义为 RNA 聚合酶结合区)，而是由多种转录因子分别特异地识别这些元件并与之结合后，RNA 聚合酶Ⅱ才能结合上去启动转录。因此，协助 RNA 聚合酶Ⅱ起作用的转录因子很多，我们可把它们分为 3 大类：

(1)基础因子(basal factor)　这些是所有启动子在启动 RNA 合成时所必需的因子(所以也叫做通用因子)。它们与 RNA 聚合酶Ⅱ结合起来形成一个围绕在起始位点周围的复合体，它们决定转录的起始位点。

(2)上游因子(upstream factor)　它们识别并特异地结合在起始位点上游的元件上，是普遍存在的，其活性不受调控，并且作用于任何具有其识别的元件的启动子。它们的作用是提高启动的效率，是那些功能相当高的启动子所必需的。每个启动子在充分表达时都需要特定的一套这种因子。

(3)诱导型因子(inducible factor)　其作用与上游因子相同，但它们是受到调控的。它们在特定的时间、条件或特定的组织中合成或被活化，因而有调控基因在不同时间、条件或不同地点表达的作用。因此，把它们结合的元件叫做应答元件(response element)。

一个仅含有被基础因子和上游因子所识别的元件的启动子能在任何种类的细胞中进行转录。可见这种启动子担负着细胞中那些组成型基因(constitutive gene)，有时也叫做持家基因(housekeeping gene)的表达。

1. 由 RNA 聚合酶Ⅱ和通用转录因子组成基础转录装置　RNA 聚合酶Ⅱ自己不能启动转录，它必须依靠转录因子的辅助。酶和这些因子一起组成基础转录装置(basal transcription apparatus)，此装置是转录任何启动子都必需的。因此，在讨论启动子的结构时，首先要确定一个通用(generic)启动子，即确定一个 RNA 聚合酶Ⅱ能在该处启动转录的最短的序列，并鉴定出识别该序列所需的转录因子。通用启动子应该能在任何细胞中启动转录，它不依赖于那些用于调控组织特异性的序列。这种启动子所需的转录因子称为基础的或者通用的转录因子，用 TFⅡX(X 代表各个因子)表示之。当然通用启动子的转录效率是很低的，通常需要有另外的上游因子才能达到应有的转录水平。

可以预料，在通用启动子中应该有保守序列，它在所有或大多数启动子中应该是同源性很高的。然而分析表明，在转录的起始处没有明显的同源序列，但 mRNA 的第一个碱基通常为 A，并且在基因中 A 的两边为嘧啶，可用通式 Py_2CAPy_5 表示之。此区称为起始子(initiator,

Inr)。仅由 Inr 组成的启动子是能被 RNA 聚合酶Ⅱ识别的最简单的启动子。

RNA 聚合酶Ⅱ的绝大多数启动子具有 TATA 框。TATA 框的共有序列为 TATAAAA，这是分析了 100 多个启动子得出的。点突变 TATA 框的个别碱基显著损伤启动子的活性，使 A 和 T 碱基对换可引起失活。可见 TATA 框是有精确序列的元件，而不只是富含 A═T 对。TATA 框位于－25 处附近，它是真核启动子中唯一与起始位点有固定距离的元件，也是最靠近起始位点的元件。少数没有 TATA 框的启动子称为无 TATA 框启动子(TATA-less promoter)。

腺病毒的晚期启动子是研究通用启动子的有用材料。它含有一个 TATA 框和一个符合 Inr 序列的起始位点，在有基础转录因子时，RNA 聚合酶Ⅱ在体外可在此启动子处启动转录。用此体系曾确定了许多启动转录所需的成分。大多数基础转录因子已被提纯，它们的基因也已克隆。

基础因子有多种，它们与 RNA 聚合酶Ⅱ共同结合在启动子处形成起始复合体。根据这些因子和 RNA 聚合酶Ⅱ结合上去的次序以及被它们保护的 DNA 区域，推测复合体形成的过程如图 13-14 所示。首先结合上去的转录因子是 TFⅡD，它结合在 TATA 框及其上游的区域上。现在已知 TFⅡD 的一个亚基是特异识别 TATA 框的，称之为 TATA 框结合蛋白。TFⅡD 的其余亚基叫做 TBP-相关因子(TBP-associated factors，TAF)。不同的 TFⅡD 可能含有不同的 TAFs，它们识别不同的启动子。这个概念还适用于其它 RNA 聚合酶的启动子。前已述及，RNA 聚合酶Ⅲ的内部启动子所用的因子是 TFⅢB，聚合酶Ⅰ启动子所用的是 SL1。可把 TFⅢB 和 SL1 看成是由 TBP 与不同套的 TAFs 结合而成的。

由于 TBP 是第一个与 DNA 结合的因子而受到特别重视。已知它是个分子量约 30 kD 的小蛋白质。它的一个很独特的性质是结合在 DNA 的小沟处(迄今已知的所有 DNA-结合蛋白都是结合在大沟)。TBP 与 DNA 结合的区域的氨基酸序列在不同品种间是保守的，而其可变的 N 端尾巴则暴露着以和其它蛋白质相互作用。提纯的 TBP 在体外与 DNA 结合时，它保护 TATA 框处约一转的 DNA 双螺旋，典型的是由－37 到－25 区。但在起始反应中 TFⅡD 的结合区为－45 到－10。TBP 是唯一与 DNA 特异序列结合的基础转录因子。

然后是 TFⅡA(酵母的由 2 个亚基组成，哺乳动物的为 3 个亚基)的结合。它的结合通过释放出一个 TAF 而使 TBP 活化。再次是 TFⅡB 的结合，它保护由－10 到＋10 区，即结合在起始位点附近。

然后是 TFⅡF 与 RNA 聚合酶Ⅱ形成的复合体结合上去。TFⅡF 由两个亚基组成。较大的亚基具有 DNA 解旋酶(helicase)活性，在启动时它可能参与解旋作用。较小的亚基能与 RNA 聚合酶Ⅱ结合得很紧。事实是 TFⅡF 与 RNA 聚合酶Ⅱ结合在一起联结到正在组装中的转录复合体上，所以 TFⅡF 有把 RNA 聚合酶Ⅱ带到复合体上去的作用。RNA 聚合酶Ⅱ的结合保护了模板链至＋15，编码链至＋20。此时 TFⅡE 能够结合了，它使保护区向下游扩至＋30。

在 TFⅡE 之后还有两个因子，即 TFⅡH 和 TFⅡJ 结合到复合体上。TFⅡH 有激酶活性，它能使 RNA 聚合酶Ⅱ的 CTD 尾巴磷酸化。这可能是 RNA 聚合酶Ⅱ与转录因子的分开所必需的，分开后 RNA 聚合酶Ⅱ便可离开启动子进行延伸。

无 TATA 框启动子的情况是怎样的？已知它需要的转录因子与有 TATA 框启动子所需要的相同，其中包括 TFⅡD。然而由于它没有 TATA 框，所以迄今尚不清楚 TFⅡD 是怎样结合上去的。看来 Inr 是个定位成分，可能有另外一个因子结合到 Inr 上，然后 TFⅡD 再结合上

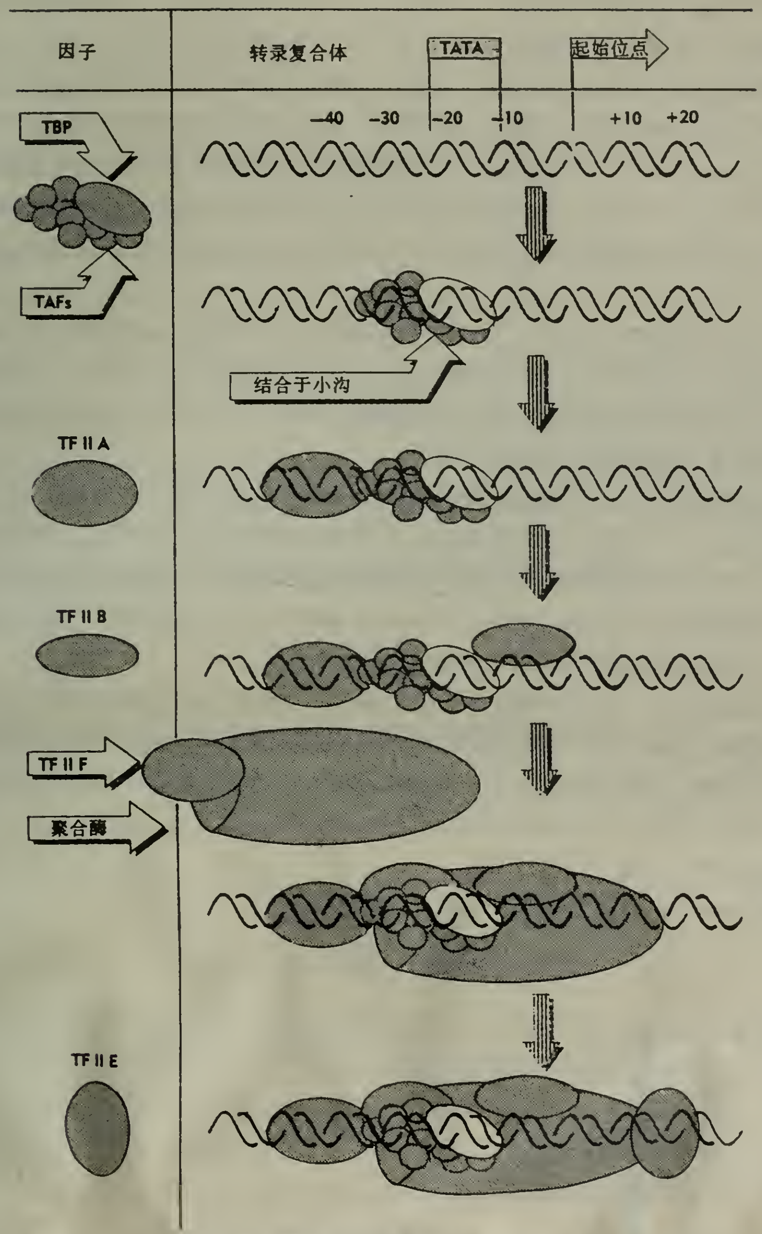


图 13-14　RNA 聚合酶 Ⅱ 的起始复合体在启动子处组装的过程

去。无论如何，TF Ⅱ D 必定是以某种与结合到 TATA 框上去的不同的方式进入复合体。可见，在无 TATA 框启动子处 TBP 的功能更像它在 RNA 聚合酶 Ⅰ 的启动子和 RNA 聚合酶 Ⅲ 的内部启动子处的作用。

许多基础因子都是由多亚基组成的，因而参与在基础转录装置中的多肽的数目很多，约有 20 多个，总分子量约为 500 kD。RNA 聚合酶 Ⅱ 本身由约 10 个亚基组成，分子量也约为 500 kD。

比较一下 RNA 聚合酶 Ⅱ 启动复合体的装配与原核 RNA 聚合酶的异同是有意义的。原核 RNA 聚合酶的核心酶不能识别启动子，要靠 σ 因子。σ 因子是 RNA 聚合酶的一个组成成分，由它识别启动子并使 RNA 聚合酶结合在启动子处。然而延伸不需要 σ 因子，所以，结合在启动子上后将之释出以便核心酶合成 RNA。因此，σ 因子起着严格规定转录起始位点的作用。看来 RNA 聚合酶 Ⅱ 的基础转录因子起着与原核中 σ 因子相类似的作用。RNA 聚合酶 Ⅱ 自己不能识别启动子，要靠基础因子来识别，所以这些基础因子起着严格规定转录起始位点的作用。但延伸不需要它们，所以在起始复合体形成之后，RNA 聚合酶 Ⅱ 要与这些因子分开，由它自己去合成 RNA。不过进化使得这些因子更独立了，它们不是组成 RNA 聚合酶的成分，而是独立的转录因子。

在有 TATA 框的启动子中，由 TATA 框调准 RNA 聚合酶 Ⅱ 的位置，使它准确地由起始位点开始转录，这大概是 TATA 框与起始位点有固定距离的原因。TATA 框的缺失引起起始位点的不稳定，尽管对转录的效率仅有较小的影响。诚然，某些无 TATA 框的启动子没有独特的起点，起点发生在一串起点中的任何一个。

2. RNA 聚合酶 Ⅱ 的上游元件及其转录因子　如上所述，RNA 聚合酶 Ⅱ 和基础因子结合在 Inr 和其附近的 TATA 框上形成基础起始复合体时是能够进行转录的，然而转录的效率很低。一个启动子要达到足够的转录效率和具有特异性，要靠其上游更远处的短序列，这些序列由上游因子或诱导型因子所识别，它们位于起点上游约 100 bp 范围内。当转录因子结合在这些序列上后，可在起始复合体形成的不同阶段发挥效应。

上游元件是通过逐个改变起始位点上游的碱基，以观察它对启动子转录效率的影响而鉴定出来的。典型的实验是用 β-珠蛋白基因进行的，如图 13-15 所示。从图中可以看出，在从起点到 −100 的范围内，大部分碱基改变不影响此启动子在 Hela 细胞中启动转录的效率。但在

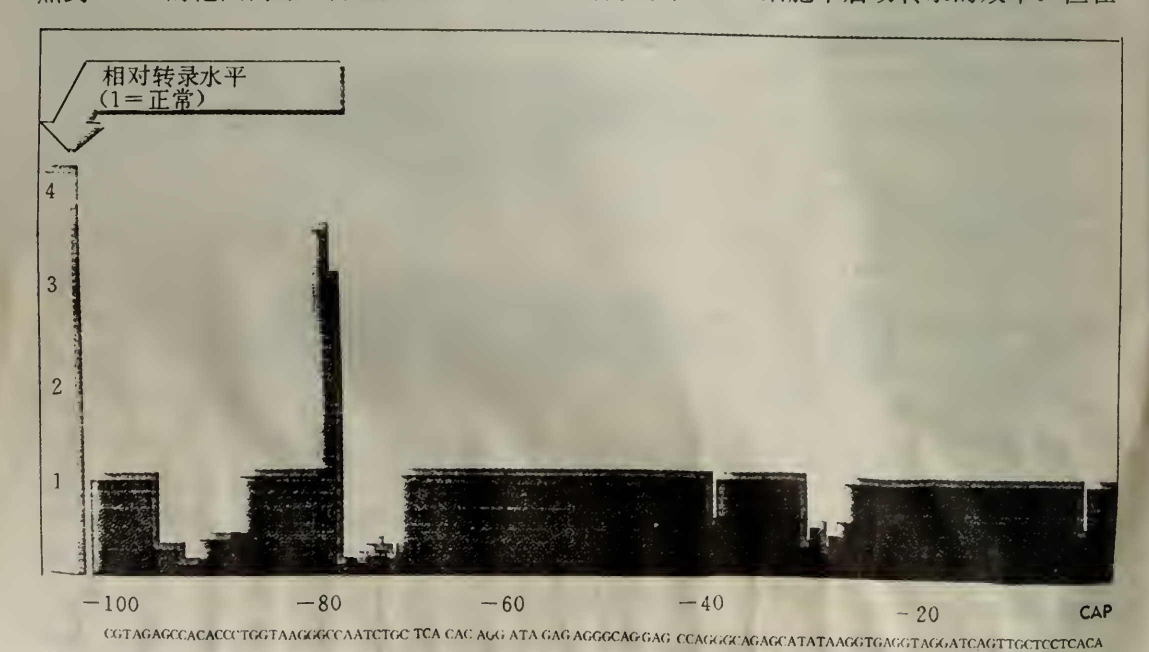


图 13-15　用碱基突变法鉴定出在 β-珠蛋白基因的上游有 3 个影响启动子转录效率的短序列，它们是 TATA，CAAT 和 GC 框

3 个短的区域内的突变使转录效率下降，它们分别集中于－30，－75 和－90 处。因而得出结论，这 3 个短的序列构成了 β-珠蛋白基因的启动子。

分析这 3 个短序列的序列得知，集中在－30 区的序列为 TATA 框。前已述及，TATA 框是规定转录起始位点的，从图中可以看出改变它的碱基比改变另外两个序列的碱基对转录效率的影响比较小。集中在－75 区的序列叫做 CAAT 框，因为它的共有序列 GGCCAATCT 中有 CAAT 而得名。集中在－90 区的序列称为 GC 框，因为其共有序列为 GGGCGG。需要注意的是，我们在写出这些元件的共有序列时，都是按编码链上的序列由 5′ 向 3′ 书写的。对于 TATA 框来说，它必须在编码链上，如果是在模板链上则无效。对于 CAAT 框和 GC 框来说，其共有序列在编码链上和模板链上同样有效，即元件的方向可以调换，但都是对其下游的基因起作用。

图 13-16 中列举了 3 个启动子作为例子。从图中可见，在这 3 个启动子中总共出现了 4 种元件，即 TATA 框、GC 框、CAAT 框和八聚体(octamer，一个 8 bp 元件)。真核启动子还有另外多种元件，它们在其它启动子中出现。从图中可以看出，每个启动子都是由特定一套多个元件所组成，但不同启动子中这些元件的种类、数目以及它们所处的位置都是不同的，而且没有一种元件是所有启动子都必须具有的，尽管有些元件可出现在多种不同的启动子中。

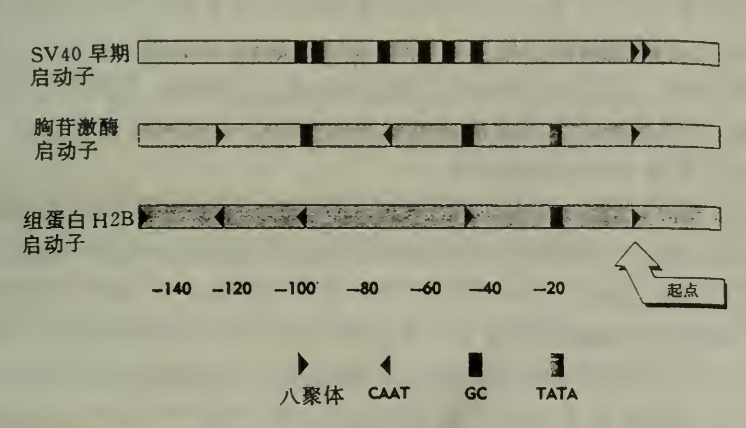


图 13-16　含有不同组合元件的几个启动子

上游元件之间的序列对启动子的功能是没有影响的。元件与元件之间的距离也是可变的，距离改变 10～30 bp 一般不影响启动子的功能。元件散布在一段很长的 DNA 上，它远远大于 RNA 聚合酶 Ⅱ 能够直接接触的范围。这提示这些元件分别由不同的转录因子所识别，这些因子结合在上游元件上后并不直接作用于 RNA 聚合酶 Ⅱ，而是作用于基础因子，再由基础因子直接作用于 RNA 聚合酶 Ⅱ。这种情况表明，在转录的启动中，蛋白质与蛋白质之间的相互作用具有和蛋白质与 DNA 之间的作用同等的重要性。

表 13-2 中列举了一些最常见的因子和它们特异识别的元件。这些因子覆盖的 DNA 序列通常比其共有序列要长些。它们通常覆盖约 20 bp DNA，而共有序列都小于 10 bp。

这些转录因子特异地结合到它所识别的元件上之后是如何活化转录的？其机制目前尚不清楚，然而已经知道了这些因子的某些一般特性。这些因子一般都是转录的激活物。典型的都有一个与特异 DNA 序列结合的特殊结构域，和一个或多个另外的结构域，这些结构域是活化转录或与其它调控蛋白相互作用所必需的。为了说明起活化作用的结构域[称为活化结构域(activating structural domain)]的某些特点，下面列举 3 种不同的转录因子作为例子，每个例

子都代表着一大类蛋白质，它们具有同一类型的活化结构域。

表 13-2 哺乳动物 RNA 聚合酶Ⅱ常用的上游转录因子及其结合的元件

元件名称	共有序列	结合 DNA 长度	因子	分子量(kD)	每细胞所含分子数	存在地点
TATA 框	TATAAAA	～10 bp	TBP	27	?	普遍存在
CAAT 框	GGCCAATCT	～22 bp	CTF/NF1	60	300 000	普遍存在
GC 框	GGGCGG	～20 bp	SP1	105	60 000	普遍存在
八聚体	ATTTGCAT	～20 bp	Oct-1	76	?	普遍存在
	ATTTGCAT	23 bp	Oct-2	52	?	淋巴
κB	GGGACTTTCC	～10 bp	NFκB	44	?	淋巴
κB	GGGACTTTCC	～10 bp	H2-TF1	?	?	普遍存在
ATF	GTGACGT	～20 bp	ATF	?	?	普遍存在

(1)SP1　它是结合到 GC 框上的转录因子。最初是在研究 SV40 病毒启动子时发现的，现在看来普遍存在于脊椎动物的细胞中。它既可与 GGGCGG 结合，也可与其互补序列CCGCCC 结合，可见 SP1 能与启动子的任一条链结合。在一个启动子中常有多个 SP1 的结合位点，例如在 SV40 的启动子中有 5 个 GC 框。SP1 是个单链蛋白质，分子量约为 105 kD。它与 DNA 结合的结构域[简称 DNA 结合域(DNA-binding domain)]在 C-端附近，含有 3 个锌指。在 SP1 中另有两个活化结构域。它们活化转录的作用机制目前尚不清楚。然而值得注意的是这些结构域中富含谷酰胺，谷酰胺约占其氨基酸残基总数的 25%。而且已知在其它与 SP1 很不相同的激活物蛋白中也有富含谷酰胺的结构域。

(2)结合 CAAT 框的因子　识别并结合在 CAAT 框上的因子有多种，其中的一类是 CTF 家族，这一类因子是由同一个基因编码，但在转录后通过不同方式的剪接产生的。CTF 成员对 CAAT 框的亲和性相同，其中之一是 CTF1。CTF1 的分子量约为 5.5 kD，它的 DNA 结合域显然没有螺旋-转角-螺旋或锌指这两种常见的蛋白质与 DNA 结合的结构模式，其结合 DNA 的机制尚未阐明。CTF1 的活化结构域富含脯氨酸，脯氨酸占其氨基酸残基总数的 20%以上。由大鼠肝中制备出来了两种结合 CAAT 框的蛋白质，即 C/EBP 和 ACF。C/EBP 优先结合 GCAAT 序列，ACF 则结合清蛋白基因启动子中的更为常见的 CCAAT 序列。C/EBP 有两个活化结构域，其中之一也富含脯氨酸，另一个不属于这里叙述的类型。C/EBP 以二聚体形式结合在 DNA 上，二聚体的两个分子以亮氨酸拉链的形式结合在一起。另一类与 CAAT 框结合的因子是 CP 家族，其成员与不同基因的启动子中 CAAT 框的亲合力不同。例如，CP1 与 α-珠蛋白基因的 CAAT 框亲和力高，而 CP2 与 γ-血纤维蛋白原基因中的 CAAT 框亲和力高。

结合 CAAT 框的因子有多种的事实提醒我们，不能认为一种元件仅被一种转录因子所识别，而且这种事实必定有其生物学意义。虽然目前我们还不清楚其意义是什么，然而看来 CAAT 框可能是个调控基因表达的靶位点。已知在组蛋白 H2B 的一个基因的启动子中有两个 CAAT 框(见图 13-17)，在海胆中此基因仅在精子发生时表达。但由睾丸和胚胎组织中都能制得结合 CAAT 框的因子，然而此因子仅能与睾丸的 CAAT 框结合，而不能与胚胎组织的 CAAT 框结合，这是因为在胚胎组织中有个叫做 CAAT-取代蛋白(CAAT-displacement protein，CDP)的蛋白质，如图 13-17 中所示，由于 CDP 结合在 CAAT 框上，从而阻止了 CAAT-结合因子的结合，使转录复合体不能形成，因而就不能进行转录了。可见 CDP 的作用类似于原核中的阻遏物。因此，结合在一个给定启动子元件上的不同蛋白质的功能可能是不同的，它们

可能是激活物、阻遏物或者具有其它功能的蛋白质。

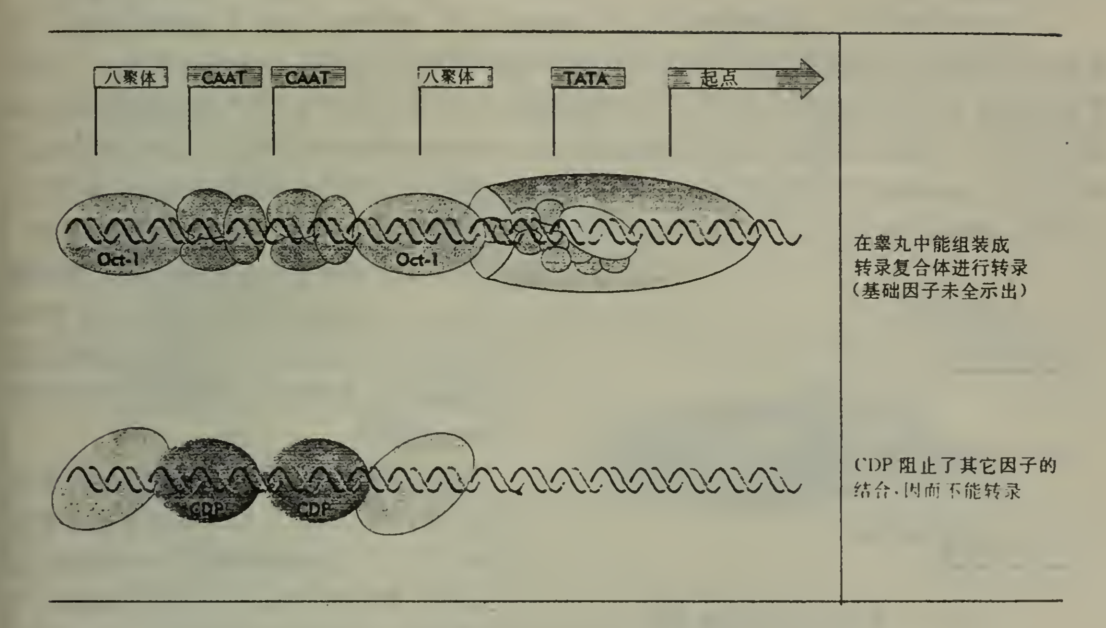


图 13-17　在海胆睾丸中的 H2B 启动子能够启动转录

但在胚胎中,由于 CDP 结合在 CAAT 框上,阻止了 CAAT 结合因子的结合,也就阻止了转录。

被不只一种因子识别的元件的另一个例子是八聚体。Oct-1 是个广泛存在的转录因子,它结合在八聚体上活化 H2B(大概还有其它)基因。在非淋巴细胞中 Oct-1 是唯一的八聚体结合因子。但在淋巴细胞中有另一个不同的因子:Oct-2,它结合在八聚体上活化免疫球蛋白 κ 轻链基因。可见,Oct-2 是组织特异性的激活物,而 Oct-1 是广泛存在的。为什么在非淋巴组织中广泛存在的 Oct-1 不能活化免疫球蛋白基因?看来八聚体前后的情况必定是重要的。Oct-2 可能是与结合在启动子上的其它蛋白质相互作用而发挥其作用,而 Oct-1 则不能。因此,我们不能简单地根据启动子中有什么元件就能预料什么因子能够活化它。

通常的情况是一个给定的共有序列被相应的转录因子所识别。但如上所述,在有些情况下,同一序列可被不同的因子所识别。而在另些情况下,一个特定的蛋白质能够识别不只一种序列。最清楚的例子是 C/EBP,如上所述它结合于 CAAT 框,但也结合于一个与 CAAT 框十分不同的序列元件上,其意义为何,尚待研究。而且我们必须考虑到,在体外观察转录因子与 DNA 结合时常常用的是裸露的 DNA。然而在细胞中 DNA 是与组蛋白结合成核小体的形式存在的,这种形式必定会对转录因子与 DNA 的结合发生影响。因此,为了研究在天然条件下活泼转录复合体的形成,必须用组装在染色质中的 DNA。总之,转录因子与启动子元件的结合是比较复杂的,其中有许多问题目前尚不明了。

(3)GAL4　是酵母中的一种激活物蛋白,它结合在一个叫做 UAS_G 的上游激活序列(upstream activating sequence,UAS)上(酵母中的上游激活序列类似于高等真核生物中的增强子),此序列在酵母半乳糖代谢有关酶类基因的附近。GAL4 的 DNA 结合域中有几个锌指,它们在 C-末端附近。其活化结构域的特点是含有许多酸性氨基酸残基。用各种肽序列取代 GAL4 的酸性活化结构域的实验结果表明,此结构域的酸性性质是关键,而其具体的氨基酸序

列则可显著改变。

3. 转录因子的DNA结合域和活化结构域是独立的　转录因子需要具有两种功能：①能够识别位于启动子、增强子或其它调节元件中的特异序列并与之结合，这些序列影响一个特定靶基因的转录；②转录因子是激活物（对于阻遏物来说，只要它结合在DNA上，就足以阻止基因的转录了），它还必须能够结合到转录装置上以发挥其促进转录的作用。实验表明，转录因子的DNA结合域和起活化转录作用的活化结构域是分别独立发挥作用的，其作用模式如图13-18所示。事实上我们可以认为，DNA结合域的功能只是把活化结构域带到转录起点的附近，而DNA结合域和活化结构域之间的联结物（connector）是足够柔性的，这样，不论DNA结合域所结合的具体位点在哪里，都能使活化结构域找到其靶蛋白，而靶蛋白很可能是基础转录因子。

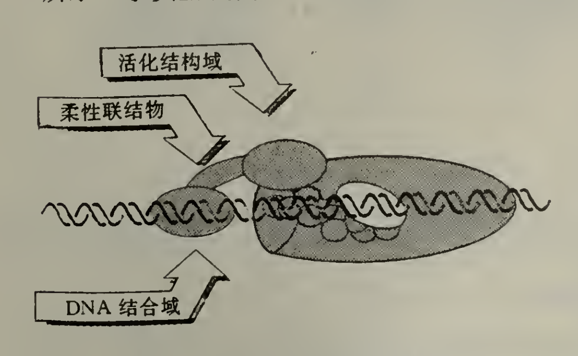


图13-18　转录因子作用的模式图

曾用酵母的激活物GAL4仔细地研究了此问题。前已述及，GAL4是结合在上游激活序列UAS_G上而发挥作用。GAL4蛋白有3个功能：即结合DNA、活化转录和结合另一个叫做GAL80的调节蛋白质。它的这些功能是可以分开的，如图13-19所示。GAL4的结合结构域由其N-端的65个氨基酸残基组成，单独的此片段能够与它的共有序列结合，但不能活化转录。GAL4有两个活化转录的区段，即148～196和768～881。如果将它们联结在DNA结合域上就能活化转录。

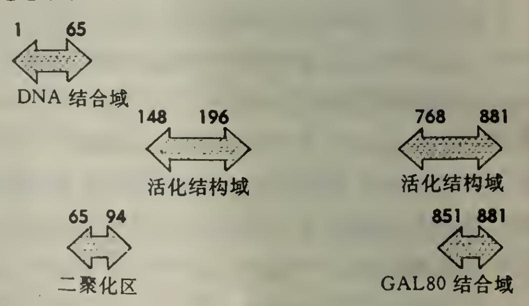


图13-19　GAL4的结构域示意图

GAL4还有另外两个功能结构域。一个是65～94区段，它可使GAL4形成二聚体。它与DNA结合域相邻，形成二聚体（与蛋白质结合）的结构域与DNA结合域相邻是在转录因子中十分普遍存在的花样。另一个是851～881区段，它位于一个活化结构域内，能与GAL80结合。后面将要谈到GAL80的结合与否是调节GAL4的活性的。

将活化结构域结合到启动子附近的DNA上当然是它能活化转录的先决条件，然而这种结合是否必须依赖其特定的DNA结合域？图13-20示出的实验结果明确地回答了此问题。LexA是细菌的一种阻遏物，其N-端有结合DNA的结构域，它特异识别LexA操纵基因。当用（通过基因工程方法）LexA的DNA结合域取代GAL4 DNA结合域时，此嵌合蛋白质不能结合在UAS_G上，因而不能活化*gal*基因。但当用LexA操纵基因序列代替酵母DNA中的UAS_G序列时，则此嵌合蛋白质能正常活化酵母基因的转录。此结果明确说明GAL4的活化结构域和DNA结合域是各自独立地发挥作用的，并且互不影响。只要能把活化结构域带到启动子附近就能发挥作用，用GAL4的DNA结合域或用LexA的DNA结合域将活化结构域结合到DNA上的效果是一样的。

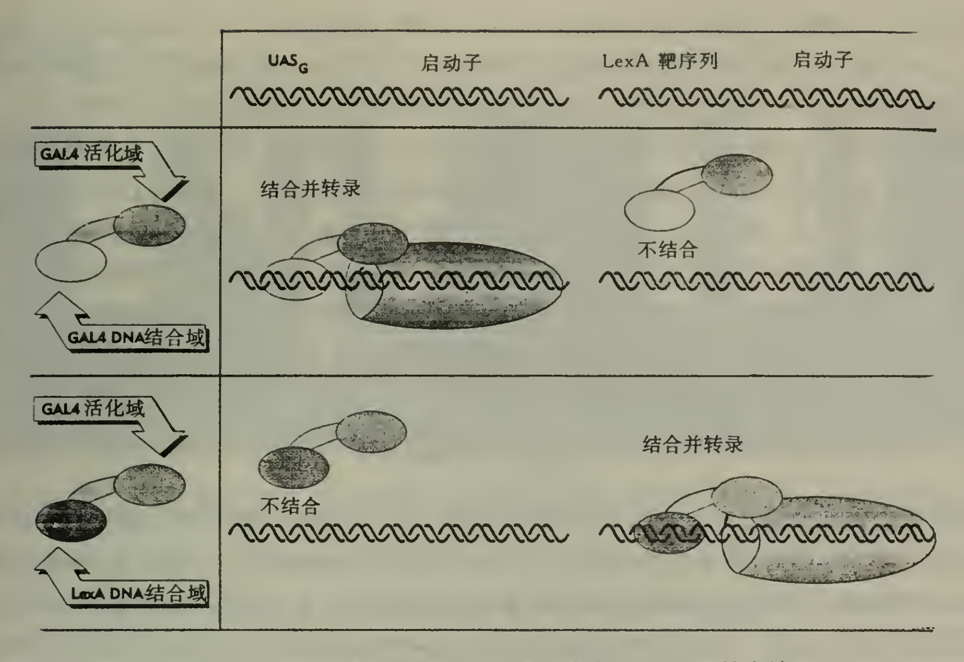


图 13-20 GAL4 活化转录不依赖于其特异的 DNA 结合域

当用 LexA 的 DNA 结合域取代 GAL4 的 DNA 结合域时，只要在启动子附近加上 LexA 的特异结合序列，此嵌合蛋白质便能活化转录。

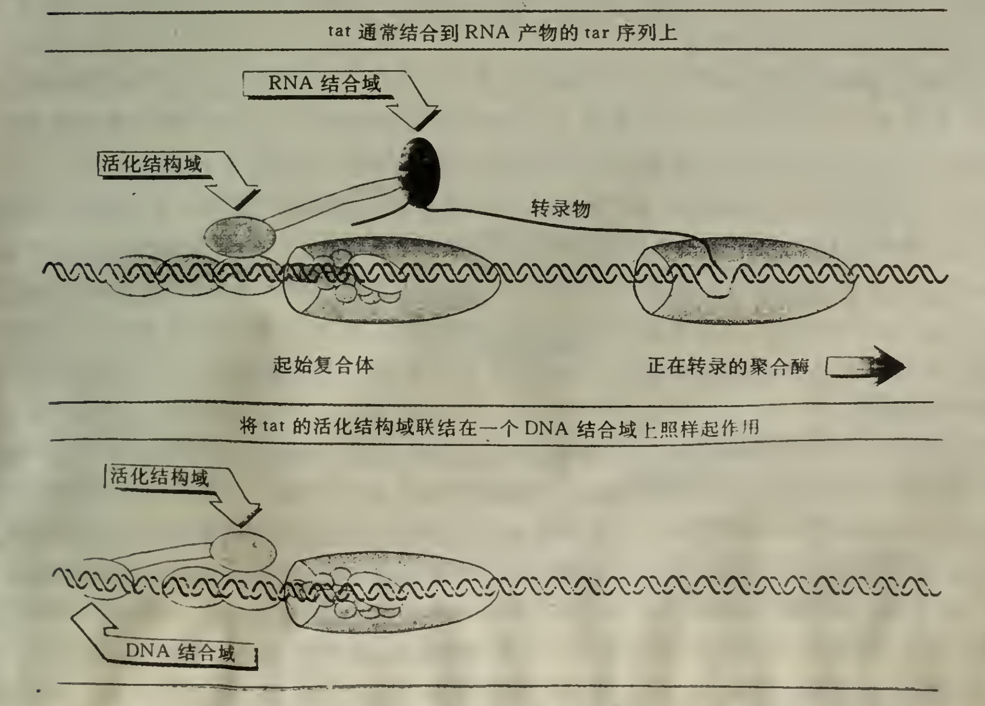


图 13-21 HIV tat 蛋白的活化结构域的作用示意图

用 HIV 的 tat 蛋白所做的实验进一步表明了转录因子的 DNA 结合域和活化结构域的作用是各自独立的。tat 蛋白不与 DNA 结合，而是结合在转录产生的 RNA 中的一个二级结构

区，此区叫做 tar 序列。如图 13-21 所示，tar 序列在转录起点的紧下游，因此，当 tat 蛋白结合在 tar 上时，就将 tat 蛋白带到转录起始复合体附近了。这样 tat 蛋白就像其它上游转录因子一样，其活化结构域便可与起始复合体中的转录因子相互作用，从而促进转录。当然，在 tat 蛋白的情况下，必须在产生第一个转录物为它提供 tar 之后，tat 蛋白才能结合上去发挥作用。

图 13-21 的下部分表明，当将 tat 蛋白改造成其活化结构域与一个结合 DNA 的结构域相联结的嵌合蛋白质时，再在启动子中加上一个该 DNA 结合域的识别序列，则此嵌合蛋白质照样能活化转录。此实验结果清楚地表明，转录因子的 DNA 结合域和活化结构域是独立发挥作用的，DNA 结合域的功能只是把活化结构域"拴在"起始复合体附近，使之能够发挥活化转录的作用而已。转录因子的这一特性表明，它们的立体结构是高度可塑的，即其 DNA 结合域和活化结构域之间的联结物的构象是可变的。

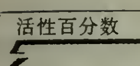
八、增强子（enhancer）

前面我们是把启动子作为担负启动转录的成分进行讨论的。然而在真核中启动子不一定是单独自己发挥作用的，至少在某些情况下启动子的活性被增强子大大地提高。增强子由另外一组元件所组成。奇异的是，增强子可在距起始位点很远（几千个碱基）的位置发挥作用，而且这种距离是可以改变的（目前还不知道距离多远后就不起作用了）。并且它们可在基因的上游、下游，甚至基因的内部发挥作用。例如，SV40 病毒的增强子是第一个发现的增强子，将它移动至此环状病毒基因组（5.2 kb）的任何位置都有作用。另外增强子序列无论在编码链或模板链上都有效。在酵母中有一种与增强子类似的序列，称为上游激活序列（UAS），但它仅在启动子上游起作用，不能在下游起作用。

前面我们曾将真核启动子定义为转录起始位点附近的一些序列，这些序列是启动转录所必需的。按此定义，启动子是不包括增强子的，因为它可距离很远。但这只是个实用的定义，并不严格，因为按其效应来说，把增强子说成为启动子的一部分，也并非不可以。

为了说明增强子的结构特点，我们可用 SV40 病毒的增强子作为例子。此增强子由两个相同的 72 bp 序列串联重复而成，已证明缺失此两个序列中的任一个都不影响转录的水平，但两者都缺失时则使转录效率大大降低。图 13-22 是用碱基突变法揭示出来的此增强子所含元件的结果。从图中可以看出，一个增强子含有多个短的序列元件，它们被特异的蛋白因子所结合。这些元件中有许多是在启动子中常见的，如 AP1 和八聚体。不过在增强子中元件的密度要比启动子中大很多。

许多增强子仅在某些细胞中有效，而在其它细胞中无效。例如，免疫球蛋白基因的增强子仅在 B 淋巴细胞中有效，从而使免疫球蛋白基因仅在 B 淋巴细胞中表达，这是因为增强子需要有特异的激活物蛋白与之结合才能促进转录，而这种蛋白质仅在某些细胞中才具有之故。在这方面了解得最清楚的是媒介糖皮质激素发挥其激素效应的增强子。糖皮质激素进入细胞后，首先与其胞液中的特异受体相互作用，这样形成的激素-受体复合体再进入细胞核，并结合在糖皮质激素的增强子上，从而引发特定的一套基因的表达，其结果表现出糖皮质激素的生理效应。可见糖皮质激素增强子仅在具有糖皮质激素受体蛋白的细胞中才能发挥作用。由此可以预料，感染真核细胞的病毒 DNA 中应该含有被寄主细胞中特异蛋白质活化的增强子。事实正是如此。例如，小鼠乳腺肿瘤病毒（MMTV）的 DNA 中含有糖皮质激素增强子，因此，MMTV 仅在那些正常受这些类固醇刺激的细胞（如乳腺上皮细胞）中才能大量繁殖，因为只有在这些



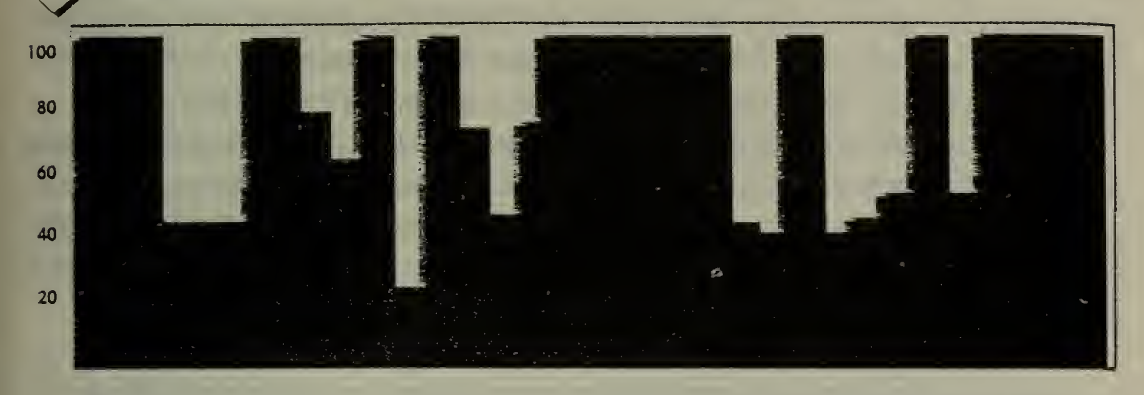


图 13-22　矩形图表示出所有使增强子的功能降至低于野生型 75％的突变，矩形图的下边示出了已知的蛋白质结合位点

细胞中才有活化其增强子的激素-受体。病毒寄主范围有局限性的原因之一，就是由于它们含有组织特异性和种属特异性的增强子之故。现在看来组织特异性表达的基因有的是由其启动子负责的，这种基因依靠其附近的增强子提高表达效率。另一些基因的启动子没有组织特异性，它们的组织特异性表达要靠增强子来调控，可见增强子具有重要的调控基因表达的作用。

增强子能在离起始位点很远处发挥作用，而且它在基因的上游或下游均可，其作用机制为何，曾经引起人们做了多种设想。现在看来，它和上游启动子元件的作用一样，是与转录的基础装置相互作用的。实验表明，当增强子未与启动子靠拢在一起时，它是不能发挥作用的。只有增强子靠拢在启动子处时才能发挥作用，这种靠拢是靠结合在 DNA 上的蛋白质起桥梁作用的，并且在增强子和启动子间的 DNA 必定发生打圈(loop)。

九、转录的调控

原则上说，真核基因的表达可在表达过程中的任一阶段进行调控。现在看来至少有 5 个潜在的调控点，即 5 种水平的调控，它们是：①基因结构的活化；②转录的启动；③转录物的加工；④mRNA 的转运到胞质及其稳定性；⑤翻译的调控。由此可见，我们不能认为，一个基因的表达一旦开始就必定要表达到底，而是不同基因的表达可在特定的环节进行调控。

前已述及，真核染色体的结构对基因的表达是有重大影响的，活泼转录的基因处的染色体结构与不转录的基因处的染色体结构不同。然而染色体结构的改变并不等于该处的基因必定被转录，而只是为转录创造了条件，即该处的基因是能够转录的。因此，基因结构的活化是基因表达的第一步，它调控着该基因的能否表达。能够被转录的基因是否真的进行转录是由转录的启动阶段调控，这就是转录水平的调控。对大多数基因来说这是主要的调控点，它大概是最普遍的调控水平。当前尚未发现在真核细胞中转录一旦启动后还会发生对转录的调控，例如在转录未完成前使之停止。

转录后的几个阶段的调控都可称之为转录后的调控。真核 RNA 聚合酶Ⅱ转录产生的初始转录物要经过多种加工后才生成成熟的 mRNA，其中包括 5′端的加帽、3′端的加尾、内含子的剪接掉和 RNA 编辑等。这些步骤都可调控，但知道最多的是在剪接阶段。成熟 mRNA 的转运入胞浆也可进行调控。最后，mRNA 在胞浆中的翻译也是重要的调控环节。在这方面对成年的体细胞了解得比较少，而在胚胎发育过程中的翻译水平的调控则了解得比较多。

为什么某些基因仅在某些细胞中表达，而在其它细胞中不表达（这是高等真核生物不同细胞具有不同表型的根本原因，是细胞分化的结果）？为什么许多基因仅在一定时间（例如发育的一定阶段）或一定条件下（包括生理的、营养的、环境的条件等）进行表达？在转录水平的调控中我们拟从以下两个方面来讨论这些问题，即：①靶基因中的什么成分能使转录因子特异地认出从而使之表达？②转录因子本身的活性是怎样调控的，而这种调控又怎样反映出各种信号的作用？

（一）应答元件使靶基因被转录因子识出

在分析一组受到共同调控的基因的结构时，发现它们都有一个相同的启动子元件，此元件被一个起调节作用的转录因子所识别。使基因对这种因子做出反应的元件叫做应答元件（response element），例如，热休克应答元件（heat shock response element，HSE）、糖皮质激素应答元件（glucocorticoid response element，GRE）和血清应答元件（serum response element，SRE）等。表 13-3 中列举了一些诱导型转录因子的性质和它们识别的元件。

表 13-3　诱导型转录因子结合到一组具有相同应答元件的启动子或增强子上使之协同调控

调节物	代号	共有序列	结合 DNA 长度	因子	分子量(kD)
热休克	HSE	CNNGAANNTCCNNG	27 bp	HSTF	93
糖皮质激素	GRE	TGGTACAAATGTTCT	20 bp	受体	94
佛波酯	TRE	TGACTCA	22 bp	AP1	39
血清	SRE	CCATATTAGG	20 bp	SRF	52

应答元件的共同特点是：它们都是启动子的上游元件或者增强子；都含有短的共有序列，并且在不同基因中应答元件的复制品（copy）很相似，但不一定完全相同；因子结合区向共有序列的任一侧扩展一个短的距离；元件与起点没有固定的距离，但通常在起点上游小于 200 bp 处；通常有一个元件即足以做出调节应答，但有时有多拷贝。

应答元件可在启动子中或增强子中。HSE 通常是在启动子中，而 GRE 则在增强子中。现在认为所有应答元件的作用原理是相同的，即一个基因由一个启动子或增强子中的序列所调控，此序列由一个特异的蛋白质所识别，此蛋白质是 RNA 聚合酶启动转录所必需的一个转录因子。仅在此基因需要表达的条件下才出现有活性的此转录因子，在没有有活性的此转录因子时意味着该启动子不被此系统所活化。

热休克应答是多个基因受单一因子调控的例子，它在原核和真核中都广泛存在。当温度升高时，某些基因关闭了，而热休克基因（heat shock genes）则被转录，并且还引起 mRNA 翻译的改变。热休克基因调控的模式在原核和真核中是不同的。在细菌中是合成了一个新的 σ 因子，它指导 RNA 聚合酶的核心酶识别热休克基因所共有的一个与普通启动子不同的 −10 区序列。在真核中，热休克基因也有个共同的共有序列（HSE），但在不同基因中它与起点的距离是可变的，并且是由一个独立的转录因子 HSTF 所识别。此因子的活化使得特异的一组（约 20 个）在启动子中含有 HSE 的基因启动转录。果蝇的所有热休克基因都含有多拷贝的 HSE，

HSTF 协同的结合在相邻的这些应答元件上。HSE 和 HSTF 在进化中都是保守的，果蝇的热休克基因可在哺乳动物或海胆中被活化。果蝇的 HSTF 与酵母的 HSTF 也很相似。在热休克时酵母的 HSTF 磷酸化了，这种修饰作用使 HSTF 活化了。

金属硫蛋白(metallothionein，MT)基因是一个基因能够受多种不同途径进行调控的例子。金属硫蛋白保护细胞对抗过浓的重金属，其方式是与金属结合并将其从细胞中移去。在一般情况下基因以一个基本的水平进行表达，但在重金属(如镉)离子或糖皮质激素的诱导下则以较高的水平表达。如图 13-23 所示，此基因的调控区有多个调控元件，在不同条件下，它们与各自的特异转录因子结合而进行调控。

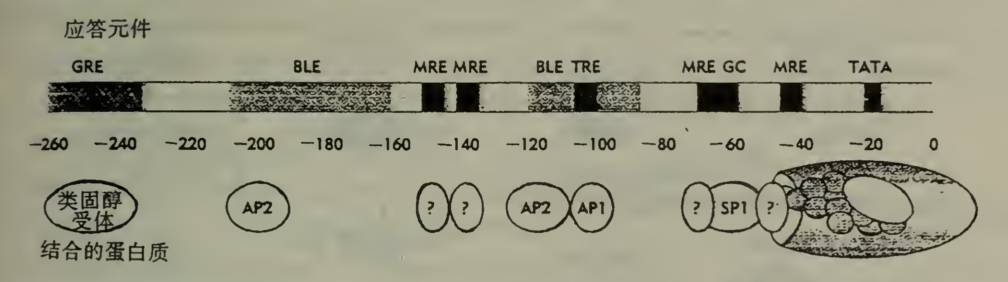


图 13-23　人金属硫蛋白基因的调控区，在其启动子中含有基础元件，并含有增强子。启动子中还含有金属诱导元件，增强子中有对糖皮质激素起反应的元件图上边示出元件，下边示出与之结合的蛋白质。

从图中可见，在起始位点附近有个 TATA 框和 GC 框，这是两个组成型的启动子元件。为了基础水平的组成型表达，除了 TATA 框和 GC 框外，还需要两个基础水平元件(basal level elements，BLE)，它们是增强子。在一个 BLE 中有个 TRE，TRE 是在数种增强子中都有的共有序列，例如，除了 MT 的 BLE 外，它还存在于 SV40 的增强子中。TRE 有个 AP1 结合位点。如前所述，AP1 是个上游因子，它的结合是组成型表达机制的一部分。然而 AP1 的结合还有第二个作用。已知 TRE 还对佛波醇(phorbol)类，如 TPA(一种促进肿瘤物质)发生应答作用，此作用是由 AP1 与 TRE 结合媒介的。

多个金属应答元件(MRE)序列担负着对金属的诱导性应答作用。它们是启动子的元件。有一个 MRE 即有对重金属发生应答作用的能力，而更高水平的诱导作用则是由多个元件达到的。对类固醇激素的应答是由一个 GRE 控制的，它位于起点上游 250 bp 处，其行为是个增强子。缺失此区域不影响基础水平的表达，也不影响金属诱导的表达水平，但它对类固醇的应答是绝对必需的。

金属硫蛋白基因的表达调控表明了一个一般的原理，即位于增强子或启动子中的几种不同元件中的任一个都能独立地活化基因。一种模式的活化作用所需的元件不存在时不影响其它模式的活化作用。元件的多样性，它们作用的独立性，以及它们的相对位置的安排显然是没有限制地可以改动，提示结合到元件上的因子是能够独立地提高由基础转录装置启动的转录效率的。之所以能够如此，大概是凭藉蛋白质-蛋白质的相互作用，以稳定或者辅助起始复合体的形成。

(二)诱导型转录因子活性的调控

基因表达调控的关键是要做到使基因在特定的时间、地点或条件下进行表达，即在需要表达时表达，不需要表达时不表达。怎样才能做到这一点? 前已述及，这些被调控的基因的启动

子或增强子中都有应答元件，当特异的诱导型转录因子（或叫做调节性转录因子，regulatory transcription factor）结合在其上时，即能启动转录。因此，调控的关键是，当需要某一（或某些）基因表达时，则提供有活性的特异转录因子，当不需要它表达时，则灭活此因子。由此可见，调控诱导型转录因子的活性是调控基因表达的关键，而诱导型转录因子活性的调控应能反映客观的需要。表 13-4 中列举了几种目前已知的调控方式。由于目前对这种调控机制的了解还不够全面和深入，因此表中所列并不全面，这种归类方式也不能作为最后定论。现将表中所列几种方式分述如下：

表 13-4　诱导型转录因子的几种活化方式

无活性情况	有活性情况	例子
合成蛋白质		
无蛋白质		同源异型蛋白
蛋白质磷酸化		
灭活蛋白质		HSTF
蛋白质脱磷酸化		
灭活蛋白质		AP1 (Jun/Fos)
配体结合		
灭活蛋白质		类固醇受体
由抑制剂释放出来		
灭活蛋白质 抑制剂		NF-κB
改换伙伴		
灭活蛋白质 灭活伙伴		HLH (MyoD/ID)

1. 调控转录因子的合成　这种调控方式的特点是，当某一基因不需要表达时，特异启动该基因转录的转录因子不存在，而需要表达时则合成此转录因子。已知的这种例子很多，突出的表现在个体发育的调控中（详见第三节）。例如，果蝇中许多具有同源异型域的蛋白质是转录因子，不同的这种转录因子分别特异地在不同类型的细胞中合成，从而使不同细胞各自表达不同

的一套基因，其结果使不同细胞具有了不同的表型。可见是由于不同细胞中存在有不同的含有同源异型域的蛋白质才使它们具有了不同的表型。事实上，不同的同源异型蛋白之所以在不同细胞中表达，是由于在不同细胞中分别具有它们各自的基因的转录因子之故。因此，这是一种基因级联的方式，即一个基因的表达产物是下一个基因的转录因子，而下一个基因的表达产物又是再下一个基因的转录因子等。这种基因级联的方式在个体发育的调控中看来很重要。

2. 通过化学修饰改变转录因子的活性　许多转录因子的活性是用化学修饰直接调控的，最常见的化学修饰是磷酸化和脱磷酸化。例如，热休克基因的应答元件(HSE)是由热休克转录因子(HSTF)特异识别的，HSTF 结合在 HSE 上启动热休克基因的表达。已知酵母的 HSTF 必须磷酸化后才有活性，在酵母热休克时，HSTF 磷酸化了，从而使得热休克基因进行表达。AP1 是通过脱磷酸化而活化的转录因子的例子。AP1 是由 Fos 和 Jun 以亮氨酸拉链方式结合起来形成的异二聚体。此转录因子仅当其 Jun 亚基脱磷酸化时才变为活性型。

3. 由配体的结合活化或灭活转录因子　对于这种调控方式知道得最清楚的是糖皮质激素受体的作用(图 13-24)。糖皮质激素类以简单扩散的方式穿过细胞膜而进入细胞。在细胞中糖皮质激素结合在其受体上。目前还不清楚游离的受体定位于何处，它们大概是以平衡状态位于核和胞质中。不管怎样，当激素结合在受体上时，受体便转变成一个有活性的转录因子。此活化的受体识别一个特异的共有序列，即糖皮质激素应答元件(GRE)。典型的 GRE 位于增强子中，此增强子与应答糖皮质激素的基因临近。因此，当激素-受体复合体结合在此种增强子上时，启动了一套基因的表达，其结果便表现出糖皮质激素的生理效应。

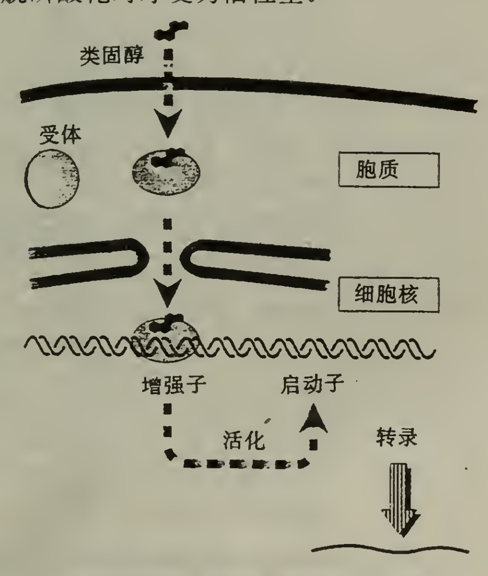


图13-24　糖皮质激素引起它们的受体与一个增强子结合来调控基因的转录

已知上述糖皮质激素的作用机制代表着一大类化合物(图 13-25)作用的共同模式。从图中可以看出，这些化合物的结构不同，有的甚至差异很大。肾上腺皮质分泌 30 多种类固醇，主要的是糖皮质激素和盐皮质激素。另两类类固醇激素(雄激素和雌激素)主要是由性腺分泌的。这些类固醇激素是应答多种神经内分泌的活动而合成并分泌的，它们对动物的生长、组织发育和体内环境恒定(homeostasis)起着重要的作用。维生素 D 则是骨骼发育所需要的。甲状腺激素不是类固醇，它调控动物的基础代谢率。类固醇和甲状腺激素在变态(metamorphosis)中可能很重要，例如，昆虫中的脱皮类固醇(ecdysteroids)和蛙的甲状腺激素。视黄酸(维生素 A)是个形态发生素，它负责正在发育的鸡肢芽体(chick limb bud)的前-后轴的发育。这些在结构和功能上都不相同的形形色色的化合物，从调控基因表达的途径来说却是用着一个相同的模式，即它们都是小分子化合物，它们各自结合在其特异的受体上，此受体活化基因的转录。

用克隆这些受体 cDNA 的方法推出了它们的氨基酸序列。分析各种类固醇激素、甲状腺激素和视黄酸的受体结构表明，它们代表着一个新的基因调节物的超家族(super-family)，即配体应答(ligand-responsive)转录因子家族。它们都有活化结构域、DNA 结合域和结合激素的

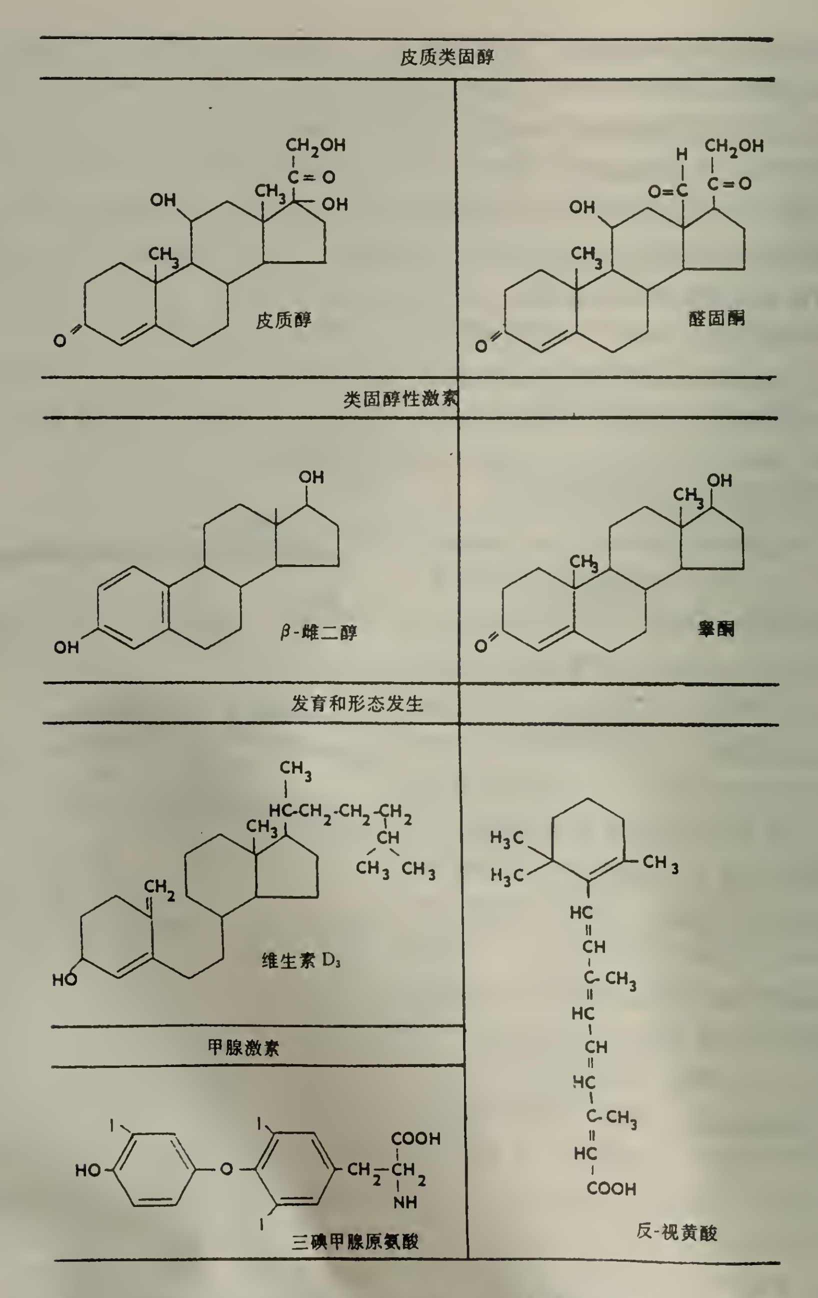


图 13-25　活化转录因子的几种小分子化合物。皮质激素和性激素是由胆固醇合成的。维生素 D 是一种类固醇。甲状腺激素由酪氨酸合成。视黄酸由维生素 A 合成

结构域，而且这些结构域的位置安排是相同的，即在受体的中心是结合 DNA 的结构域，此区域在这些受体中是比较保守的。它在活化转录中也起作用，因为当此结构域发生突变时，既不能与DNA 结合，又丧失了受体活化转录的作用，即活化结构域和DNA 结合域不是独立的。受体的 N-端含有起活化转录作用的另一部分成分。C 端是与激素结合的结构域，此结构域与DNA 结合域是各自独立的。

但不同受体的调节方式并不完全相同。例如，糖皮质激素受体的C-端结构域缺失时，缺失后的蛋白质是组成型的活化基因的转录，即不再需要类固醇来活化它。这提示在完整受体中，当没有类固醇激素时，其结合激素的结构域阻止了受体识别GRE，即它起着内在的负调控作用。糖皮质激素的结合使此阻遏作用失活，受体能结合GRE并活化转录。当前对此阻遏作用的机制尚未确定。一种可能是受体自己内部的作用；另一种可能是有某种另外的蛋白质参与，此种蛋白质与受体结合时阻遏了受体与GRE的结合，而具有糖皮质激素时，激素取代了此种蛋白质而使受体活化。

在雌激素受体中各结构域之间的相互作用与上述糖皮质激素的受体不同。当结合激素的结构域缺失时，雌激素受体的剩余部分仍能与ERE结合，但不能活化转录。可见，在活化转录中需要其结合激素的区域，但具体的机制尚不明了。

在类固醇受体家族的成员中C-端结构域的同源性在30%～57%范围内，它们与甲状腺激素等的受体的同源性相比则小得多，表明了此结构域与不同化合物结合的特异性。

这些受体以二聚体形式与其DNA上的应答元件相结合，结合的方式是锌指，例如，糖皮质激素受体和雌激素受体都有两个锌指。这些受体的应答元件都是由两个短的重复序列所组成，每个重复序列也叫做半位点(half site)。两个半位点之间以0～4 bp分隔开，它们或是回文形式，或是直接重复。正是用这种间隔不同和重复方式不同将不同受体的应答元件区别开的，而半位点的共有序列种类则很有限。已知糖皮质类固醇受体(GR)、盐皮质类固醇受体(MR)、雄激素受体(AR)和孕酮受体(PR)都形成同二聚体，它们识别的应答元件半位点的共有序列为TGTTCT。两个半位点安排为回文，位点间的间隔决定着元件的类型。雌激素受体(ER)的作用方式与上述相同，但其半位点的序列为TGACCT。

甲状腺激素受体(T3R)，维生素D受体(VDR)，视黄酸受体(RAR)和9-顺视黄酸受体(RXR)则以异二聚体形式发挥作用，它们识别的半位点的序列为TGACCT，但两个半位点的安排为直接重复。由于两个半位点之间的间隔不同而被不同的受体所识别，其中RXR的应答元件两个半位点间的间隔为1 bp，VDR的为3 bp，T3R的为4 bp，RAR的为5 bp。

4. 抑制因子的解除　例如，NFκB在B淋巴细胞中活化免疫球蛋白κ基因的转录因子，但它也存在于多种其它细胞中。在其它细胞中由于抑制因子I-κB(一种蛋白质)与NFκB相结合，使NFκB滞留在胞浆中，不能发挥作用。在B淋巴细胞中，则NFκB由I-κB的结合中释放出来，并移动至核中活化转录。

另一个例子是酵母中的GAL4。GAL4结合在上游活化序列UAS_G上时，启动UAS_G附近的一些基因转录，这些基因的表达产物是半乳糖代谢的有关酶类。但当没有半乳糖时，蛋白质GAL80与GAL4相结合，使GAL4没有启动转录的活性。在有半乳糖时，则半乳糖促使GAL4-GAL80复合体解离，产生有活性的GAL4，启动有关的酶类基因的转录。这是个典型的反馈控制的例子，实际上是半乳糖自己诱导产生代谢它的酶类。

5. 二聚体伙伴的改变　这种方式的调控可用螺旋-环-螺旋(HLH)蛋白作为例子来说明。在前面曾经谈到，HLH蛋白有碱性的(bHLH)和非碱性的两类。在一个二聚体中，如果两个亚基都是bHLH蛋白，则它能够和DNA特异结合。如果两个亚基中有一个是非碱性HLH蛋白，则此二聚体不能与DNA特异结合。这就是这种调控方式的基本原理。下面举两个具体的例子(见图13-26)。

bHLH 蛋白质二聚化并与 DNA 结合 | 非碱性的 HLH 蛋白质阻止与 DNA 结合

HLH
碱性区
MyoD/E12
或
AC-S/da
HLH
无碱性区
E12/Id
或
AC-S/emc
不能与 DNA 结合

图 13-26　在一个 HLH 二聚体中，如果两个亚基都是 bHLH，则能与 DNA 结合。如果二聚体中的一个亚基没有碱性区，则不能与 DNA 结合

在果蝇中，基因 *emc*(extramacrochaetae)是建立正常成年感觉器官的立体样式所必需的，它是通过它的表达产物抑制几种基因产物的功能而发挥作用的，其中包括基因 *da*(daughterless)和 achaete-scute complex(*AC-S*)产物的功能。如不抑制这些基因产物的功能，则它们将引起额外的细胞形成感觉器官。*AC-S* 和 *da* 的表达产物为 bHLH 类蛋白质。*emc* 编码的是没有碱性区的 HLH 蛋白，它是个抑制物。可以设想，在没有 *emc* 功能时，Da 和 AC-S 蛋白质形成二聚体，这些二聚体活化适当的靶基因进行转录。但当产生 Emc 蛋白时，则引起 Emc 蛋白与它们形成二聚体，此种二聚体不能与 DNA 结合，也就不能活化靶基因的转录。可见，在适当细胞中产生 Emc 蛋白以抑制 *AC-S/da* 的功能，是使这些细胞不形成感觉器官所必需的。

肌肉细胞的生成是由转录程序的一个改变触发的，此程序需要数种 bHLH 蛋白质，其中包括 MyoD。MyoD 是在生肌细胞(myogenic cells)中特异产生的，如果在某些其它细胞中过度表达 MyoD，则能引起这些细胞开始一个生肌程序。肌肉分化的触发物大概是由 MyoD-E12 或 MyoD-E47 形成的异二聚体，而不是 MyoD 的同二聚体。在生肌作用开始之前，一个非碱性 HLH 蛋白，Id 蛋白与 MyoD 和/或 E12，E47 相结合，形成一种不能与 DNA 结合的异二聚体。Id 与 E12，E47 结合的能力比与 MyoD 强，可见 Id 的作用是束缚住广泛表达的 E12 或 E47，而使它们不能发挥作用。Id 的过度表达能阻止生肌作用。可见移去 Id 是触发 MyoD(与 E12，E47)启动生肌作用的机制。

上述两个例子表明 HLH 蛋白类在细胞分化中起着重要的作用。两种 bHLH 特异结合形成的二聚体调控着特定基因的转录，从而使细胞发生特定的分化。而非碱性 HLH 的存在抑制着两种 bHLH 二聚体的形成。可见，细胞的分化与否依赖于特定的非碱性 HLH 蛋白的移去或存在。

十、mRNA 前体的可变剪接

和原核不同，真核编码蛋白质的基因经转录产生的初始转录物需经过加工后，才产生出成熟的 mRNA，然后 mRNA 进入胞浆，经翻译产生出特定的蛋白质。mRNA 前体的这种加工包括 5′加帽、3′加尾、剪接和编辑等。现在已知许多这些加工过程是受到调控的。例如，通过不同的剪接可由一个基因的转录物产生出不同的成熟 mRNA，从而翻译出不同的蛋白质。又如，通过编辑后 mRNA 的密码子有所改变，产生出与基因编码有所不同的蛋白质等。这些都具有重要的生物学意义。下面先介绍 mRNA 前体的可变剪接，然后再介绍 RNA 的编辑。

（一）mRNA 前体的可变剪接的概念

大多数真核基因转录产生的 mRNA 前体是按一种方式剪接产生出一种 mRNA，因而只产生一种蛋白质。但有些基因产生的 mRNA 前体可按不同的方式剪接，即可变剪接(alternative splicing)，产生出两种或更多种 mRNA。在有些情况下，虽然由于可变剪接而产生的 mRNA不同，但是由于其差异仅在 5′或 3′非翻译区，因而产生的蛋白质是相同的。这种可变剪接的意义如何，目前尚不甚明了，但推测它可能对翻译有调控作用。在另些情况下，则由于可变剪接所产生的不同 mRNA 的编码区有所不同，因而产生出不同的蛋白质。其中有的产生两种蛋白质，有的可产生更多种。在这些例子中，有些是多种蛋白质在同一细胞中同时产生；有的则是同一基因在不同细胞中产生出不同的蛋白质，表现出组织特异性；有的则是在不同发育时期或条件下采取不同的剪接方式，表达出不同的蛋白质，所有这些都是受到调控的。

为什么真核基因是割裂的，它含有那么多内含子？这是迄今仍不甚明了的问题。但可变剪接提供了一种事实，可用以说明内含子的功能之一，即由于内含子的存在而使得可变剪接成为可能。其意义在于可使一个基因表达出多种蛋白质，即扩大了 DNA 中遗传信息的含量。然而令人不解的是，高等真核生物的基因组是很宽阔的，它完全能够容纳更多的基因，但不知为什么一方面它的许多基因非常分散，另方面它又使一个基因产生出多种产物。

（二）可变剪接的方式

现在已知许多基因有两个或更多个启动子，因而有两个或更多个转录起始位点。当用不同的启动子进行转录时产生的初始转录物是不同的。然而迄今尚未发现由于所用启动子不同而产生出不同蛋白质的例子。目前尚不清楚一个基因有多个启动子的意义如何，但估计可能与基因表达的调控有关。现在还发现许多基因的初始转录物有不止一个加尾位点（图 13-27 A）。与选用不同启动子的情况不同，选用不同的加尾位点可产生出不同的蛋白质（如免疫球蛋白）。另一种可变剪接方式是选用不同的剪接位点（图 13-27 B），如SV40 的 T 和 t 抗原是由同一基因用这种方式产生出来的两种蛋白质。再一种方式是上述二者兼有，如降钙素基因转录产物的剪接方式。

（三）可变剪接的举例

1. 在大鼠中，降钙素(calcitonin)基因的初始转录物（图 13-28）有 6 个外显子和 2 个加尾位点　在甲状腺中加尾时采用第一个位点，使加尾后的产物仅含有前 4 个外显子（其中第四个为降钙素所特有），它再经剪接产生成熟的 mRNA，此 mRNA 的翻译产物为降钙素。而在脑中加尾时采用了第二个加尾位点，因而加尾后的产物中含有全部 6 个外显子。但在剪接时将第四个外显子剪接掉，使成熟 mRNA 由 1，2，3，5，6 等 5 个外显子组成，其翻译产物为另一种激素，叫做降钙素基因相关肽(calcitonin gene related peptide，CGRP)。这个例子说明，同一基因的转录产物在不同组织中进行了不同的剪接，从而产生出不同的蛋白质，它必然是受到调控的。

在这个例子中还可看出，它既采用不同的加尾位点，又进行不同的剪接。

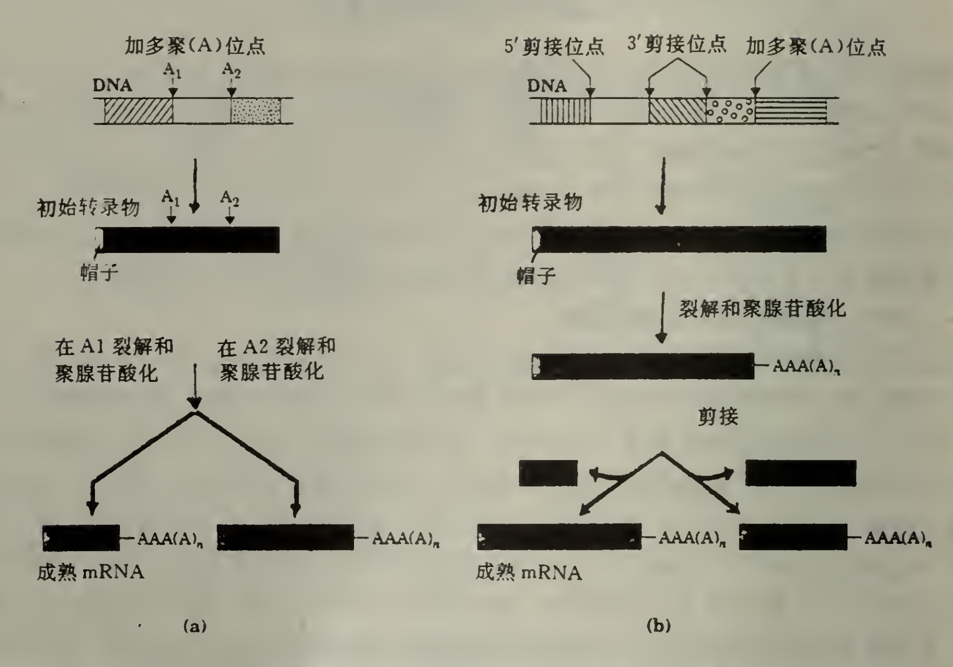


图 13-27　初始转录物不同加工过程的两种机制

(a)多个加尾位点(图中示出2个，A_1 和 A_2)的可变使用；(b)可变剪接(示出2个不同的3′剪接位点)。

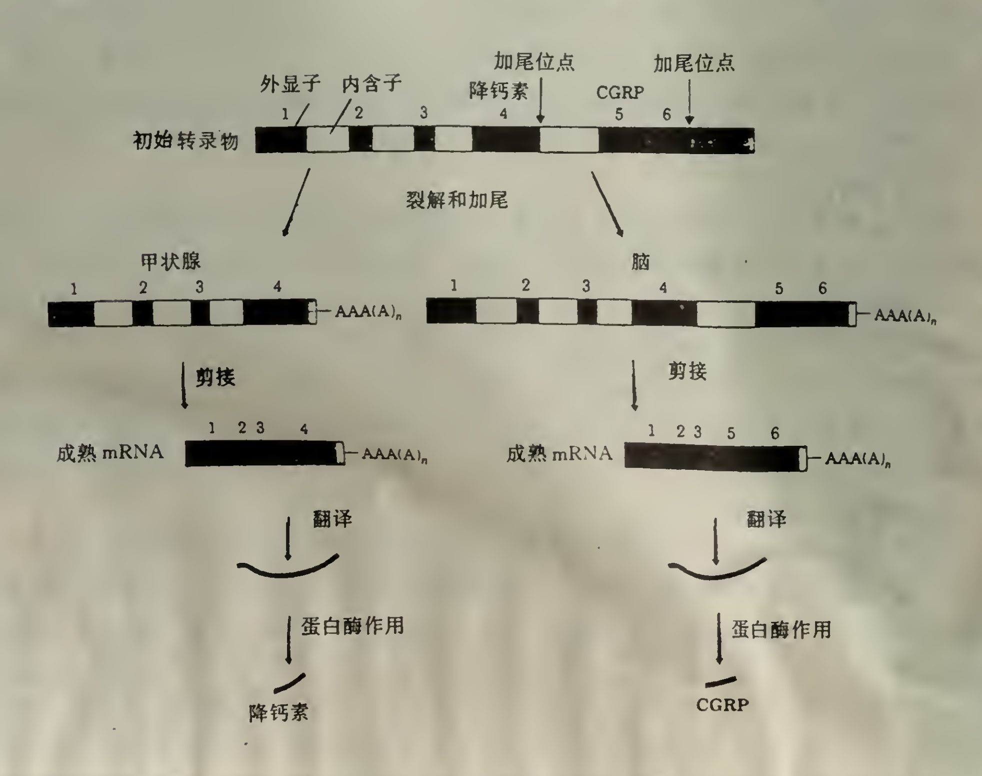


图 13-28　降钙素基因转录产物在不同组织中进行不同的加尾和剪接，从而产生出不同的蛋白质

2. 在产生抗体的细胞前体，前 B 淋巴细胞(pre-B lymphocyte)中合成的抗体是定位在细胞质膜上的，此种抗体的重链的 C 端有段疏水区，此疏水区嵌在膜中，从而把抗体固定在质膜上(图 13-29)　当用特异的抗原刺激前 B 淋巴细胞时，抗原被细胞膜上的抗体所识别，引起细胞的分化和增殖。这种活化了的 B 淋巴细胞产生的抗体被分泌到胞外，因为它的重链的 C 端没有那一段定位在质膜上的疏水区。

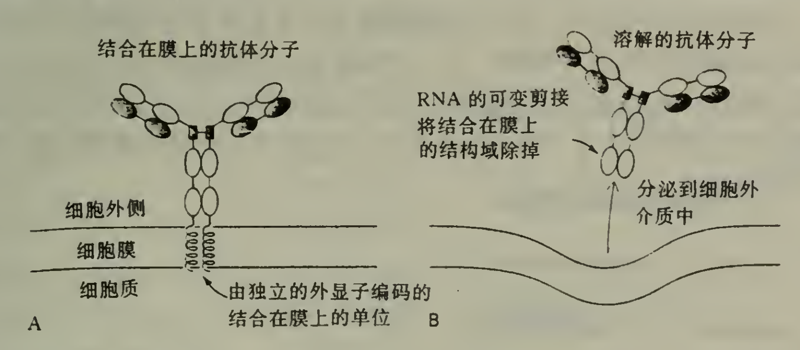


图 13-29　A. 前 B 淋巴细胞产生的抗体定位于膜上；B. 活化的 B 淋巴细胞产生的抗体被分泌到胞外

考察发生这种变化的原因得知是由于可变剪接引起的。图 13-30 中所示是 IgM(免疫球蛋白 M)的情况。由图中可以看出，IgM 重链的 mRNA 前体在恒区(constant region)有 6 个外显子和 2 个加尾位点。在前 B 淋巴细胞阶段，加尾发生在外显子 M2 之后，因而其恒区是由剪接在一起的 6 个外显子产生的。前 4 个外显子编码恒区的 4 个结构域。后 2 个外显子(M1 和 M2)编码 41 个氨基酸残基的疏水的 C-末端区以及一段不翻译的尾序。M1 和 M2 编码的一段疏水区就是使 IgM 定位在质膜上的序列，因而此时产生的 IgM 是定位在质膜上的。当前 B 淋巴细胞受到抗原的刺激而变为活化的 B 细胞，即分泌阶段时，mRNA 前体的加尾发生在外显子 $\mu 4$ 之后，因而剪接后的成熟 mRNA 中没有了外显子 M1 和 M2，产生的 IgM 便是分泌型的了。这是一个仅因选用的加尾位点不同而产生不同蛋白质的例子，蛋白质间的差异仅在 C 端。

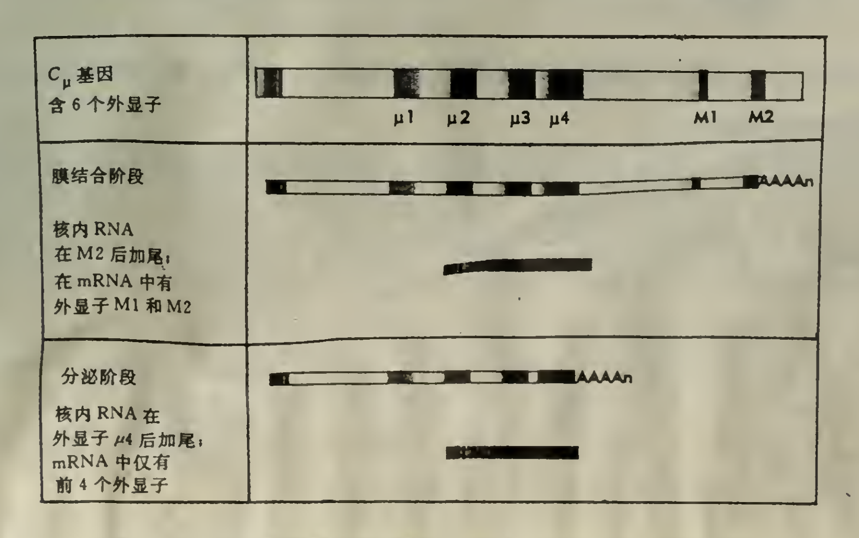


图 13-30　IgM 重链基因转录物的不同加工方式，仅示出恒区部分

这个例子说明，在细胞发育的不同阶段，调控机制可使同一基因的转录产物发生不同的剪接，产生出不同的蛋白质，以和不同发育阶段的需要相适应。

3. SV40 病毒的早期基因转录物通过不同方式的剪接产生两种 mRNA，它们的翻译产物分别为 T 抗原(分子量约为 100 kD)和 t 抗原(分子量约为 18 kD)，如图 13-31 所示 SV40 早期基因的转录物含有两个外显子和一个内含子，它有两种剪接方式。一种是剪接掉一个大的内含子，产生一个比较小的 mRNA。由于在剪接掉的内含子中有个终止密码子也被除去了，因而翻译出来的是分子量比较大的 T 抗原。另一种剪接方式是剪接掉一个比较小的内含子，产生一个比较大的 mRNA，但是由于在其中保留了那个终止密码子，所以翻译产生的是分子量小得多的 t 抗原。这种剪接方式是 3′剪接位点固定，但选用了不同的 5′剪接位点。两种剪接方式在同一细胞中同时发生。

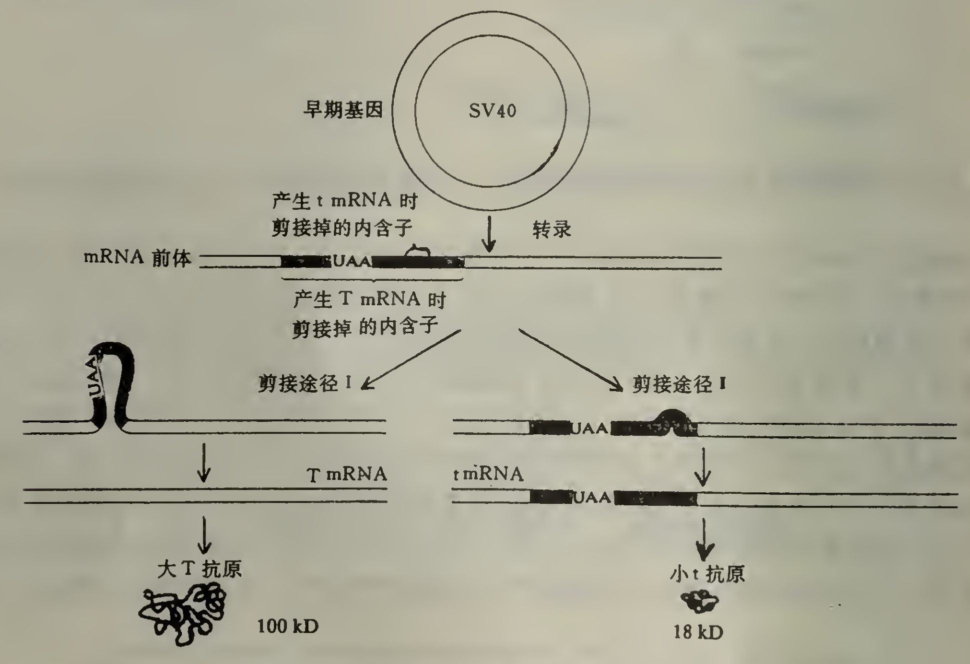


图 13-31 SV40 的早期转录物通过不同的剪接表达出 T 抗原和 t 抗原两种蛋白质

(四)可变剪接的调控机制

当前迫切需要解决的问题是什么物质调控着可变剪接？其调控机制如何？对于这些问题目前已经有了一些了解，但还远不彻底。从前面所举的 SV40 的例子可以看出，在剪接时是用相同的 3′剪接位点与不同的 5′剪接位点相联的，从而产生了 T 抗原和 t 抗原两种蛋白质。在产生 T 抗原的剪接中，由于选用的 5′剪接位点去掉了一个终止密码子，所以产生一个大的蛋白质。而在产生 t 抗原的剪接中采用了另一个 5′剪接位点，因而在其 mRNA 中保留了该终止密码子，产生出一个较小的蛋白质。在腺病毒科(Adenovirus)的 E1A 区的转录产物剪接时，也是用同一 3′剪接位点与不同的 5′剪接位点联结，从而产生出 3 种蛋白质(图 13-32)。而在果蝇的 tra 剪接中则是用同一 5′剪接位点与不同的 3′剪接位点联结，结果由于一个终止密码子的保留与否而引起了不同的结果，从而决定了果蝇的性别(详见后)。

不同剪接位点的选用是怎样做到的？在用SV40感染不同类型的细胞时发现它们产生的T/t抗原的比值是不同的。用产生较多t抗原的细胞抽提物在体外进行实验时，其剪接方式也是产生较多的t抗原mRNA。由这种类型的细胞中提取到一种蛋白质，当将此蛋白质加在其它类型细胞的抽提物中时，也使其它类型细胞的抽提物产生出较多的t抗原mRNA，说明此蛋白质具有可变剪接的效用，并曾经把它叫做可变剪接因子(alternative splicing factor, ASF)。后来发现它原来就是剪接因子SF2。SF2是剪接体组装的早期步骤以及剪接中第一个裂解-连接反应所必需的因子，它是个与RNA结合的蛋白质，具有一个结合RNA的富含精-丝二肽序列的基元。此基元在其它结合RNA的蛋白质中也具有。当一个mRNA前体的一个3′剪接位点前面有不止一个5′剪接位点时，提高SF2的浓度促使它使用与3′位点最近的5′位点进行剪接，而减少用其它5′位点。SF2的这种效应能被另一个剪接因子SF5所抵消。这些因子的确切分子作用目前尚不清楚，但总的说来，参与在剪接体组装中的蛋白质影响着5′剪接位点的选择性使用。

SV40的T/t抗原是分别用不同的5′位点与一个3′位点剪接

t

T

腺病毒E1A用不同的5′位点与一个3′位点剪接

1 2 3 4 外显子

13S 289氨基酸

12S 243氨基酸

9S 55氨基酸

果蝇的tra用一个5′位点与不同的3′位点剪接

雄和雌均无蛋白产生

只有雌性产生一个200氨基酸的蛋白质

图13-32 在剪接时选用不同的剪接位点可引入终止密码子(用*号示出)或改变读框，从而产生出不同的蛋白质

（五）可变剪接决定果蝇的性别

在果蝇的性别决定途径中涉及到一系列基因间的相互作用，在其中用可变剪接区分开雄和雌。图13-33示出了此途径，在其中X染色体与常染色体的比值决定着Sxl(Sex-lethal)的表达，并且表达的改变逐次地通过其它基因到达Dsx(double Sex)，Dsx是此途径的最后成员。

途径是由Sxl的性别特异性剪接开始的。*sxl*基因的外显子3含有一个终止密码子，它阻止功能蛋白质的合成。在雄性(X染色体对常染色体的比值低)剪接的mRNA中含有此外显子，因而不能产生出功能的蛋白质。在雌性(X染色体对常染色体的比值高)剪接的mRNA中不含此外显子。因而仅在雌性中产生Sxl蛋白。Sxl蛋白具有与其它结合RNA的蛋白质类似的结构。

Sxl蛋白的存在改变了transformer(*tra*)基因的剪接(可变剪接的方式与*dsx*类似，参看图13-32)。一种剪接方式是在雄性和雌性中都发生的，剪接产生的mRNA中含有靠前的终止密码子，因而没有功能蛋白质的产生。在雌性中由于Sxl蛋白的存在，它阻止了与5′位点靠近的3′剪接位点的利用，而是越过它用了下一个3′剪接位点，从而去掉了一个外显子，也就去掉了其中的终止密码子。由此产生了一个雌性特异的mRNA，它编码一个蛋白质。因此*tra*仅在雌性中产生一种蛋白质，此蛋白质是个剪接调节剂，它影响*tra2*的剪接，以类似的方式使得

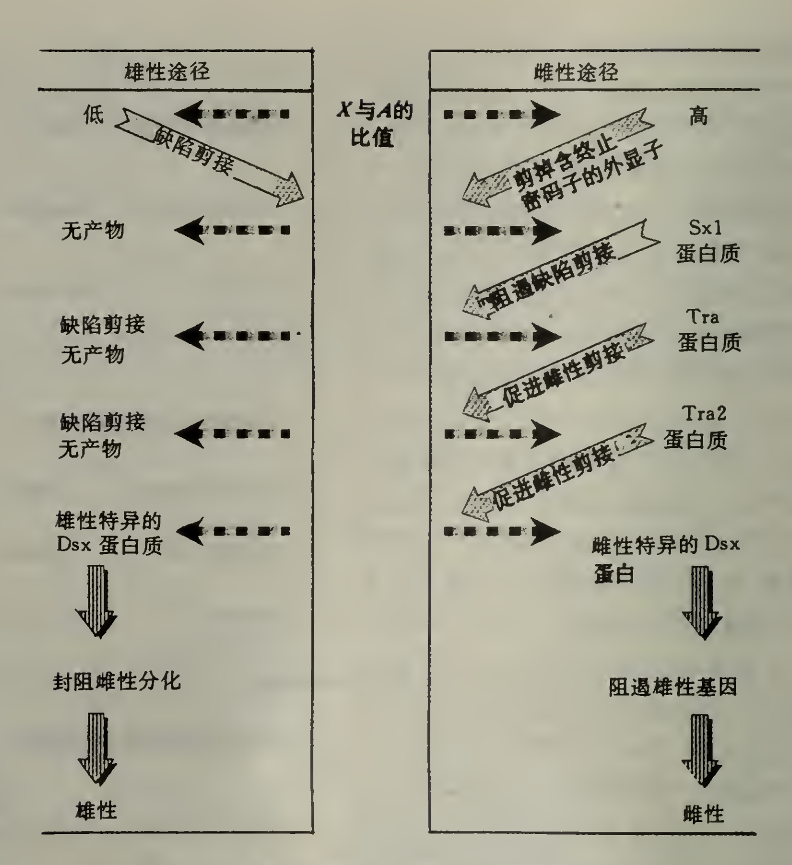


图 13-33　果蝇(*D. melanogaster*)性别决定的途径

在雌性中发生一系列不同的剪接，在每一步都封阻了雄性发育的途径。

仅在雌性中产生出 Tra2 蛋白。Tra2 蛋白再影响 *dsx* 的剪接。Tra 和 Tra2 与 SF2 类似，都有精-丝样子的结合 RNA 的基元，它们直接作用于靶转录物。*tra*2 的雌特异性剪接需要 Tra，而 *dsx* 的雌性特异性剪接需要 Tra 和 Tra2 二者。

图 13-34 表示出 *dsx* 的可变剪接方式。从图中可以看出，果蝇的 *dsx* 基因在雌性和雄性中的剪接方式不同。在雌性中(有 Tra 和 Tra2)是内含子 3 的 5′剪接位点与其 3′剪接位点相接，从而保留了外显子 4。外显子 4 中有个终止密码子，所以翻译至此为止，产生的是雌性特异的 Dsx 蛋白。在雄性中(没有 Tra 和 Tra2)内含子 3 的 5′剪接位点直接与内含子 4 的 3′剪接位点相剪接，使其 mRNA 中没有了外显子 4，因而翻译可继续到外显子 6，产生的是雄性特异的 Dsx 蛋白。可见在此途径的最后在雌性和雄性中产生出不同的蛋白质。雄性的产物封阻了分化为雌性，而雌性的产物则阻遏了雄性特异基因的表达。

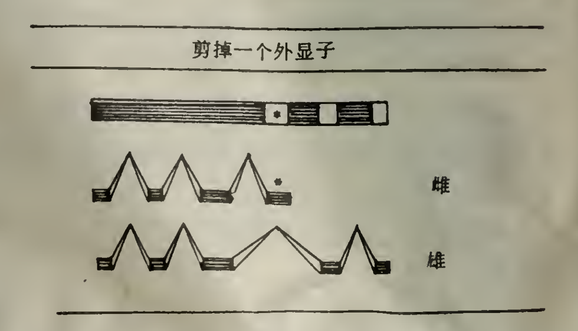


图 13-34　果蝇 *dsx* 基因转录物在雌性和雄性中的不同剪接方式

注意外显子 4 中有个终止密码子(以 * 号示出)。

综上可见，性别决定起始于 *sxl* 在雌性中的特异剪接，它避免了像雄性中那样的缺陷性剪接，从而仅在雌性中产生 Sxl。Sxl 影响 *tra* 的剪接，Tra 又影响 *tra2* 的剪接，从而使得仅在雌性中产生 Tra 和 Tra2，它们再影响 *dsx* 的剪接，最后决定了性别。已知 Tra2 在体外结合在 *dsx* mRNA 前体中的雌性特异外显子内的一个短区域上，并诱导雌性特异剪接，但仍不清楚它是怎样调控 3′剪接位点的选择的。由于没有 Tra 和 Tra2 时能够发育为正常的雄性，可见正常剪接不需要它们。它们作为特异的调节物，不必须参与在剪接反应的机制之中。在这一点上它们与 SF2 不同，SF2 是一般剪接所需要的因子，同时也影响不同剪接位点的选择。

（五）反式剪接

前面所讨论的剪接过程都是发生在一个 RNA 分子内部，即通过剪接将一个 RNA 分子内部的内含子剪接掉，使外显子连接在一起，这种剪接方式称为顺式剪接(cis-splicing)。已发现，在锥虫、线虫以及植物的叶绿体中发生两个 RNA 分子之间的剪接，即反式剪接(trans-splicing)。因此，在真核生物中虽然反式剪接比顺式剪接少很多，但它确实存在。

在锥虫(*trypanosome*)中发现有许多 mRNA 的 5′端有个共同的 35 碱基的前导序列(leader sequence)。然而此前导序列并不被每个转录单位的上游所编码，而是在由位于基因组其它区域的一个重复单位转录出来的独立的 RNA 中。此 RNA 的 3′端还具有前导序列以外的一段序列。图 13-35 中示出了此 RNA 的组成。在其 35 碱基前导序列的紧后面是个有 5′剪接位点的序列。而在编码 mRNA 的序列中，在成熟 mRNA 的紧前面有个 3′剪接位点。

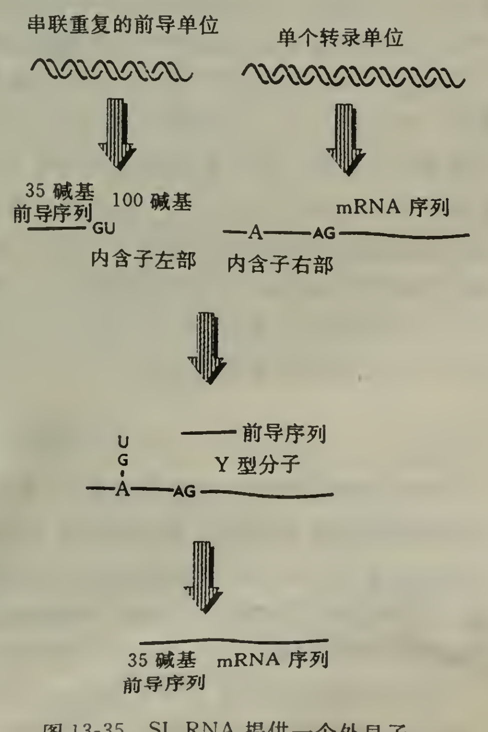


图 13-35 SL RNA 提供一个外显子，此外显子通过反式剪接与一个 mRNA 的第一个外显子相连接

如图 13-35 中所示，如果前导序列与 mRNA 是由反式剪接连接起来的，则前导 RNA 的 3′区和 mRNA 的 5′区各相当于一个内含子的一半。如果剪接是按通常的核内剪接机制进行的，则首先应在内含子 5′段的 GU 与 3′段 AG 上游的分支点之间发生反应，形成一个 5′-2′连接的磷酸二酯键。此时由于内含子的两个部分不是共价连接在一起的，因而形成的是个 Y 字型分子，而不是像通常那样形成个套索型分子。已证明此种剪接确实是形成个 Y 字型分子，因而支持图 13-35 中所示的剪接模式。

线虫(*C. elegans*)中肌动蛋白(actin)基因的表达与上述情况类似。3 种肌动蛋白 mRNA(以及某些其它 RNA)的 5′端是个相同的 22 碱基的前导序列。此序列不在肌动蛋白基因中编码，而是由另外的基因编码，此基因独立地转录出一个 100 碱基的 RNA，前导序列是此 RNA 的一部分。

在反式剪接中提供 5′外显子的 RNA 称为剪接的前导 RNA (spliced leader RNA, SL RNA)。SLRNA存在于几种锥虫中，也存在于线虫中。它们有一些共同的特点，它们都折叠成

一种相同的二级结构，此结构有 3 个茎-环(stem-loop)和一个单链区，是以 snRNA 的形式存在的。锥虫具有 U2、U4 和 U6 snRNAs，但没有 U1 或 U5 snRNAs。可用SL RNA 的性质来解释为什么没有 U1，因为 SL RNA 能够完成通常是由 U1 snRNA 完成的在 5′剪接位点的功能。可见，SL RNA 事实上是含有一个具有 U1 功能的 snRNA 序列，此序列与它识别的外显子-内含子位点联结在一起。看来，SL RNA 的反式剪接反应可能代表着向完全的核内剪接装置进化中的一个步骤。

某些叶绿体基因是反式剪接的，图 13-36 中指出了一个例子。单胞藻(*Chlamydomonas*)叶绿体的 *psa* 基因由 3 个相隔很远的外显子所组成。外显子 1 和外显子 2 相距 50 kb，外显子 2 和外显子 3 相距 90 kb。在这些外显子之间有许多其它的基因。而且外显子 1 的转录方向与外显子 2 和 3 的转录方向相反。因此，无论如何它们也不可能是被转录出一个 RNA 分子进行表达的。事实上已鉴定出来的转录物只是含有每个外显子(和它两侧的局部序列)的 RNA。因而推断其 mRNA 是由两个反式剪接反应形成的。在这两个反式剪接反应中还需要一些其它的基因，因而可知此 mRNA 的剪接过程是很复杂的。

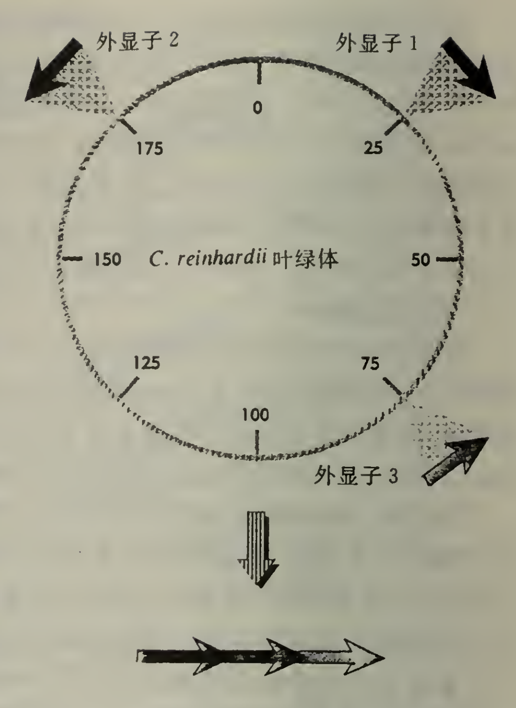


图 13-36　叶绿体 *psa* 基因的外显子分布情况。推测其 mRNA 是由两步反式剪接形成的

十一、RNA 编辑(RNA editing)

分子生物学的一个基本公理是，一个 mRNA 的序列只能代表 DNA 中编码它的序列。中心法则设想的是个线型关系，即 DNA 的一个连续的序列转录为 mRNA 序列，mRNA 序列再直接翻译成蛋白质。割裂基因的存在以及通过 RNA 的剪接移去内含子，在基因表达的过程中引入了一个额外的步骤，即 DNA 中的编码序列(外显子)必须在 RNA 中再连接。但是此过程仍然是个信息传递过程，在其中 DNA 的真正编码序列没有发生改变。

在一些特殊的条件下 DNA 编码的信息发生了改变，其中最显著的是在哺乳类和禽类中产生编码免疫球蛋白的新序列。这些改变特异地发生在合成免疫球蛋白的体细胞(B 淋巴细胞)中。在再造一个免疫球蛋白基因的过程中，在单个细胞的 DNA 中产生新的信息，并且在 DNA 中编码的信息由于体细胞的突变而改变了。不过在这种情况下，DNA 中的信息仍被忠实地转录到 RNA 中。

与上述情况不同，RNA 编辑是个信息在 mRNA 水平上发生改变的过程。此过程是由于一个 RNA 中的编码序列与转录产生它的 DNA 序列有所不同而揭示出来的。已知的 RNA 编辑有两种：①在哺乳动物细胞中发生的 mRNA 个别碱基的取代，从而引起它编码的蛋白质的改变；②在锥虫的线粒体中一些基因的转录物发生更多的改变，产生碱基的插入或缺失。

图 13-37 所示是哺乳动物中载脂蛋白 B(apo-lipoprotein B，apoB)mRNA 编辑的过程。

apoB 包围在脂质的外面形成脂蛋白颗粒，因而在甘油三酯和胆固醇的转运中起着重要的作用。apoB 有两种。在肝中合成的是分子量约为 512 kD 的 apoB-100，它参与内生性脂质的运输；另一种是在小肠中合成的分子量约为 240 kD 的 apoB-48，它以乳糜微粒的形式运输由食物中吸收的脂肪。apoB-100 含有 4 563 个氨基酸残基，apoB-48 由 apoB-100 的 N 端 2 152 个氨基酸残基所组成。由于 apoB-48 没有了 apoB-100 的 C 端部分，所以它虽然能形成脂蛋白颗粒，但不能像 apoB-100 那样与细胞表面的 LDL（低密度脂蛋白）受体结合了。

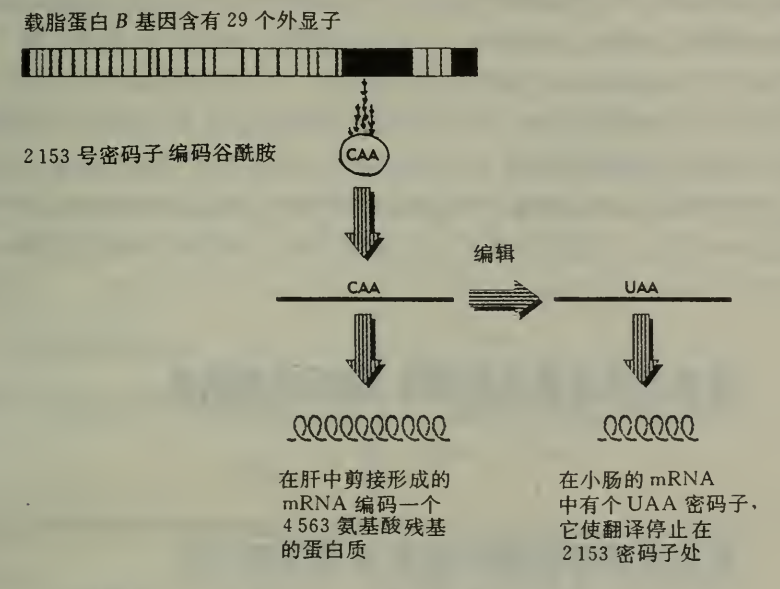


图 13-37　小肠和肝中 *apoB* 基因的序列是相同的，但在小肠中 mRNA 的一个碱基改变了，从而产生了一个终止密码子

已知在基因组中仅含有一个编码 *apoB* 的基因，并且在所有组织中都是相同的。此基因的编码区有 4 563 个密码子，在肝中它转录出一个 mRNA，其编码序列全部被翻译而产生肝中的 apoB-100。那么为什么在小肠中会产生出一个短了约一半的蛋白质 apoB-48？经查明，小肠中的 mRNA 与肝中的 mRNA 有所不同，有一个碱基发生了改变，即在肝的 mRNA 中第 2 153 号密码子为谷酰胺的密码子 CAA，这与 DNA 中的是一致的。但在小肠中此密码子中的 C 变成了 U，成为终止密码子 UAA，所以在小肠中的翻译到此为止，产生出一个比肝中小得多的蛋白质。此现象在人、兔、鼠中都已发现。

此种 C 变为 U 的过程是怎样发生的？在基因组中没有编码此种新序列的另外的基因或外显子，也未发现在剪接方式上有任何改变。因而认为这是一种新的机制在起作用，是在转录产生 RNA 的过程中或在转录完成后发生 C→U 的转变，即在 RNA 水平上发生的，称之为 RNA 编辑。已知 apoB mRNA 中的 C 变为 U 是由特异的脱氨基酶催化的，此酶仅存在于小肠中，肝中没有，它是受发育调控的。有实验表明，在 apoB 的编辑中编辑位点下游 5～15 一段核苷酸序列很重要，称为“系留（mooring）区”。如将此 11 个核苷酸用其它核苷酸替换，则引起编辑活性丧失；如将此段序列移到其它位置或其它 mRNA 上，则会产生新的编辑位点。因而推测此序列可能是蛋白质识别并结合的区域，特异的脱氨基酶以某种方式到达此位置，并催化编辑位点的胞嘧啶脱去氨基变成尿嘧啶。

在哺乳动物细胞中用 RNA 编辑的方式在 RNA 水平上改变信息的事实虽然很少，但 apoB 的编辑不是唯一的。另一个已知的例子是小鼠脑中的谷氨酸受体。谷氨酸结合在突触后(postsynaptic)膜上的受体上时开放脊椎动物中枢神经系统中的阳离子特异通道(cation-specific channels)。RNA 编辑将谷氨酸受体 mRNA 中的一个谷酰胺密码子(CAG)改变成为精氨酸的密码子(CGG)。受体中的谷酰胺变为精氨酸阻止了 Ca^{2+} 通过此通道，但 Na^{+} 可以通过。可见编辑作用是生理功能所必需的。这种 A 变为 G 的编辑机制目前尚不甚明了。

已发现锥虫线粒体的一些基因的序列发生了很大的改变。第一个发现的是其细胞色素氧化酶亚基Ⅱ蛋白的序列与编码它的基因 *cox* Ⅱ 相比有移码(frameshift)现象。图 13-38 示出了其基因和蛋白质的序列。这种情况是怎样发生的？从图中可以看出是由于在 *cox* Ⅱ mRNA 的移码位点附近插入了 4 个额外的尿苷酸。由于这种插入才恢复了该蛋白质的正确读框(翻译出正确的功能蛋白质)。插入的结果是增加了一个氨基酸，并且改变了周围的氨基酸。由于未能发现与其 mRNA 序列相同的另外的基因，因而认为这些额外的碱基是在转录过程中或转录后插入的。与此类似的 mRNA 与基因组序列之间的差异在一些病毒中也已发现，在其中是在 mRNA 中插入 G。

I S S L G I K V E N L V G V M 基因组中编码的

A TATCAAGT T TAGGTATAAAAGTAGA G A ACC TGGTAGGTGTAAT DNA 序列

−1 移码

AUAUCAAGUUUAGGUAUAAAAGUAGA UUG UA UACCUGGUAGGUGUAAU RNA 序列

I S S L G I K V D C I P G R C N 蛋白质序列

图 13-38　锥虫 *cox* Ⅱ 基因的 mRNA 与其 DNA 相比有个 −1 移码
正确的读框是经插入 4 个尿苷酸产生的

类似的 RNA 序列的编辑在其它基因中也发生，并且包括有尿苷酸的插入和缺失。图 13-39 是布氏锥虫(*T. brucei*) *cox* Ⅲ 基因的编辑情况。在此很极端的例子中，mRNA 中一半以上的尿苷酸是不被基因编码的。将基因组 DNA 与 mRNA 进行对比可以看出，没有长于 7 个核苷酸的一段序列是没有改变的，而且有的地方连续插入了 7 个尿苷酸之多。

T

UAUAUGUUUUGUUGUUUAUUAUGUGAUUAUGGUUUUGUUUUUUA UUGG UAUUUUUUAGAUUUA

UUUAAUUUGUUGAUAAAUACAUUUUAUUUGUUUGUUAA UUUUUUUG UUUUG UGUUUU UGGUU

TTT　TTTT

UAGG UUUUUUU G UUG UUG UUGUUUUGUAUUAUGAUUGAGUUUGUUGUUUGGUUUUUUGUUUU

TTT　TTTT

UUG UGAAACCAG UUAUGAGAGUUUGCAUUG UUAUUUAUUACAUUAAGUUGG UGUUUUUGG UUC

图 13-39　*T. brucei cox* Ⅲ　mRNA 的部分序列
指出了许多未被 DNA 编码的 U(下面划线者)和由 RNA 移去的 U(以 T 表示出)。

特异插入 U 的信息是由什么提供的？已知一个指导 RNA(guide RNA)的序列是与正确编辑后的相应 mRNA 片段互补的。图 13-40 示出了指导序列在利什曼原虫(*Leishmania*)的细

胞色素 b 基因中的作用模型。上面的序列是初始转录物（或者叫做编辑前的 RNA，pre-edited RNA）的序列。其中的间隔表明在编辑过程中将在该处插入碱基。在此区域必须插入 8 个 U 才能产生正确的 mRNA 序列。指导 RNA 序列与 mRNA 的编辑区及其周围的相当长的一段是互补的。典型的情况是在编辑区的 3′侧互补的长一些，而 5′侧则比较短。在指导 RNA 与编辑前的 RNA 配对时出现间隔，这些间隔是指导 RNA 中的 A 未能配对处，因为在编辑前的 RNA 中该处没有与之配对的 U。这样指导 RNA 提供了一个模板，使得在这些位置能够插入缺少的 U。编辑完成后，指导 RNA 便与 mRNA 分开。编辑的程序可能是很复杂的。在此例子中，在很长的一段转录物中要通过编辑总共插入 39 个 U 残基（图中未全示出），看来它需要两个指导 RNA，它们作用在相邻的位置。编辑可能是同时进行，也可能是先用一个指导 RNA 进行编辑后再用另一个指导 RNA 进行编辑。已在 *Leishmania* 的线粒体 DNA 中测定出来了细胞色素 b 的基因，它编码编辑前的 mRNA 序列；另有两个区域是编码指导 RNA 的。这两个编码指导 RNA 的区域以及细胞色素 b 基因在基因组中是由其它基因将它们分散开的。所以指导 RNA 是由独立的转录单位编码的。

基因组 AAAGCGGAGAGAAAAGAAA A G G C TTTAACTTCAGGTTGTTTATTACGAGTATATGG

转录

编辑前的 RNA AAAGCGGAGAGAAAAGAAA A G G C UUUAACUUCAGGUUGUUUAUUACGAGUAUAUGG

与指导 RNA 配对

编辑前的 RNA AAAGCGGAGAGAAAAGAAA A G G C UUUAACUUCAGGUUGUUUAUUACGAGUAUAUGG

指导 RNA AUAUUCAAUAAUAAAUUUUAAAUAUAAUAGAAAAUUGAAGUUCAGUAUACACUAUAAUAAUAAU

插入尿苷酸

mRNA AAAGCGGAGAGAAAAGAAAUUUAUGUUGUCUUUUAACUUCAGGUUGUUUAUUACGAGUAUAUGG

指导 RNA AUAUUCAAUAAUAAAUUUUAAAUAUAAUAGAAAAUUGAAGUUCAGUAUACACUAUAAUAAUAAU

释出 mRNA

mRNA AAAGCGGAGAGAAAAGAAAUUUAUGUUGUCUUUUAACUUCAGGUUGUUUAUUACGAGUAUAUGG

图 13-40 编辑前的 RNA 中编辑区的两侧与一个指导 RNA 配对，指导 RNA 作为插入 U 的模板。插入后产生的 mRNA 与指导 RNA 互补

原则上说，突变其“基因”或者它的任何一个指导 RNA 都能改变 mRNA 的序列，因而改变蛋白质。按照基因的标准，可把这些转录单位的每一个都看成是“基因”的组成部分。由于这些单位是独立表达的，所以它们当然是以反式（trans）补充（complement）的。由此可见，一个蛋白质的一级序列需要由 3 个相互补充的单位编码。

部分编辑的中间产物的鉴定提示编辑作用是沿编辑前的 RNA 的 3′→5′方向进行的。指导 RNA 不仅藉助与编辑前的 RNA 配对以规定 U 的特异性插入，而且它还是插入 U 的给体。指导 RNA 转录终止在一段 3～5 个 U，然后再加上一个 3′寡聚 U 尾巴，寡聚 U 的长度为 5～24 个 U 残基。这些 U 被插入到编辑的 RNA 中，其反应与 RNA 自我剪接类似。指导 RNA 与

编辑前的 RNA 之间配对后，通过两个转酯化作用将 U 插入，如图 13-41 所示。在第一步反应中，和编辑前的 RNA 配对的指导 RNA 的 3′-OH 攻击编辑区，使编辑前的 RNA 裂解为两部分。一部分是其 5′区段，它有个 3′-OH；另一部分是其 3′区段，它的 5′端与指导 RNA 的 3′端共价连接接在一起。第二步反应是第一步反应产生的编辑前的 RNA 5′区段用其 3′-OH 攻击指导 RNA 中两个 U 之间的磷酸二酯键，这样再产生编辑前的 RNA，但已在其中插入了一个 U。同时再生成了指导 RNA，但它已丢失了一个 U。重复此过程，便可在编辑位点处插入多个 U。有实验表明每次仅插入一个 U，而不是插入一组。看来其过程可能是在每加上一个 U 后，便用指导 RNA 的互补性检查一下是否还需要插入。如果需要则再插入一个，如果不需要就不进行了，即正确的编辑序列是逐步完成的。

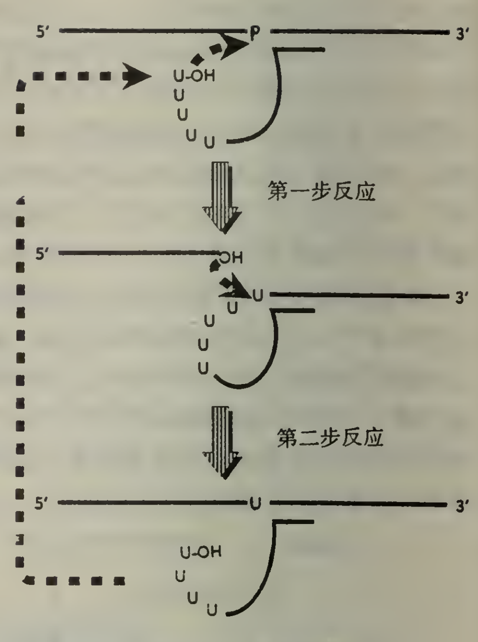


图 13-41　插入 U 的两步转酯化作用

除了上述例子之外还发现了一些其它编辑的事实。例如，已发现植物线粒体中的细胞色素氧化酶亚基、细胞色素 b 和 NADH 脱氢酶等的 mRNA 都是经过编辑的，其序列与它们基因组中的相应 DNA 序列有所不同。已知的编辑方式是用一个 U 代替一个 C。但与 apoB 的情况不同，植物线粒体中的这种代替不是引进一个终止密码子，而是改变一个氨基酸。此外，许多编辑的机制目前还不清楚。例如，有的编辑是缺失 U，有的是插入 G 等，这些都需要研究。

和可变剪接一样令人不解的是，真核生物为什么要采取编辑的方式！

十二、mRNA 稳定性的调控

和原核的相比（原核 mRNA 的寿命通常为几分钟），真核 mRNA 的寿命比较长。典型真核 mRNA 的半衰期通常为 8～10 h。然而有些真核 mRNA 的寿命却比较短，例如，调控细胞周期的蛋白质的 mRNA 的半衰期为 10～30 min，这是与这些蛋白质的功能相适合的。调控细胞周期的蛋白质是在细胞周期的特定时相中合成的，此时相一过它的合成便需停止。因此，编码这些蛋白质的基因（已知的有致癌基因、细胞分裂素基因等）是短暂表达的，它们的表达要与细胞的分裂过程同步，随着细胞分裂的过程而及时表达和迅速关闭。这类蛋白质的合成能够迅速停止的原因之一，就是由于它们的 mRNA 的寿命比较短。

研究这类 mRNA 的寿命比较短的原因发现在它们的 3′端非翻译区都有个相似的序列，其形式为 AUUUA。已证明此序列是 mRNA 在细胞中不稳定的原因之一。缺失掉此序列可使原来不稳定的 mRNA 变为稳定，而且在本来稳定的 mRNA 中加入 AUUUA 序列后，可使之变为不稳定。为什么有此序列的 mRNA 不稳定？这当然与 mRNA 的降解机制有关，目前正在积极探讨有关 mRNA 降解的机制。

已知有些 mRNA 的稳定性是受到调控的，并从而调节它编码的蛋白质的合成。其中知道得比较清楚的是运铁蛋白受体（transferrin receptor，TfR）mRNA。TfR 含量的多少调节着细

胞摄入的铁量，而 TfR 合成的速度（因而其含量）则由细胞内铁的水平所调控。当细胞内铁的含量下降时，TfR 的合成增强，从而摄入更多的铁，使细胞内铁的水平上升至符合需要的量。而铁的含量多时则降低 TfR 的合成速度。这种保持细胞内铁含量恒定的反馈控制是怎样做到的？已证明 TfR mRNA 稳定性的调控是调控 TfR 合成速度的重要环节，而 TfR mRNA 的稳定性则是受到其 3′端非翻译区的一段特异序列所调控。此段序列可形成 5 个茎-环（stem-loop）结构（见图 13-42）。已知一个叫做 IRE-结合蛋白的蛋白质结合在此茎-环结构上时，使 TfR mRNA 稳定，不结合此蛋白质的 TfR mRNA 则易于被降解，当细胞内铁缺乏时，IRE-结合蛋白结合在 TfR mRNA 上，使之稳定，从而合成出比较多的运铁蛋白受体，细胞因而摄入较多的铁，当细胞内铁丰足时，则 IRE-结合蛋白不结合在 TfR mRNA 上，使 TfR mRNA 易于降解，合成的 TfR 便较少，细胞摄入的铁就少了。一个与此有关的机制还调控着铁蛋白的合成，将在下面讨论。

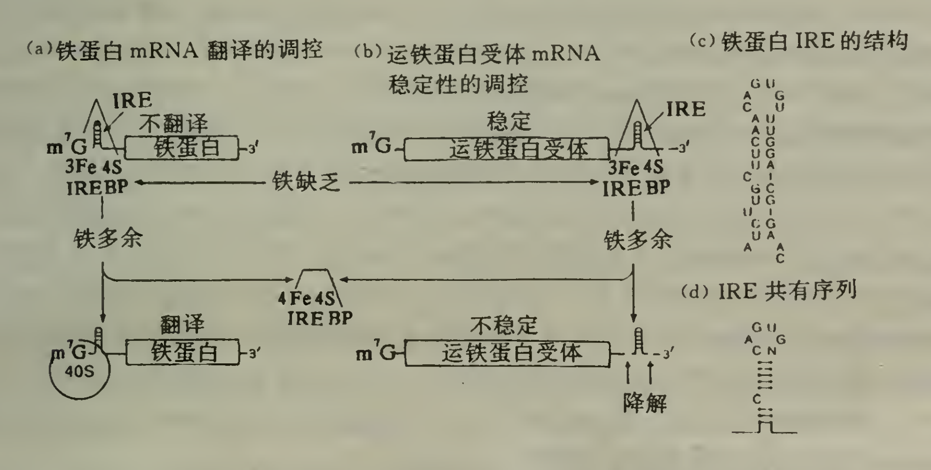


图 13-42　哺乳动物细胞中铁蛋白基因和运铁蛋白受体基因转录后调控示意图

十三、翻译的调控

在真核基因表达的翻译调控中，目前知道得比较清楚的有 3 个例子：①铁应答的 mRNA-结合蛋白对铁蛋白 mRNA 翻译的抑制作用；②通过真核起始因子-2 调节翻译；③氨基酸缺乏对酵母 GCN4 翻译的诱导作用。目前已知的真核翻译调控的例子比原核的少，但看来这不是由于翻译的调控对真核不重要，而是由于研究真核翻译的调控比原核难得多。相信在经过深入研究之后，将揭示出更多的真核翻译调控的例子和规律。下面着重介绍这 3 个例子，并提供一些有关真核翻译调控的已知线索。

（一）铁蛋白合成的调控——阻遏物蛋白对特异 mRNA 的调控

哺乳动物细胞中的铁蛋白（ferritin）是贮存铁的蛋白质，它在细胞中的含量是随细胞内铁含量的变化而变化的。即在细胞内铁的含量低时，铁蛋白的合成速度减慢，使细胞内铁蛋白的含量下降；而在细胞内铁的含量高时，则铁蛋白的合成速度加快，使细胞内铁蛋白的含量升高。可见，是细胞内铁的水平调控着铁蛋白合成的速度。这种调控是怎样实现的？

已知这是在翻译水平上进行的调控，是由一个约 90 kD 的蛋白质，它结合在铁蛋白mRNA 的 5′-末端附近抑制翻译的进行（图 13-42 a）。此阻遏物结合的靶位点是个保守的茎-环结构，

如图 13-42 c 和 d 所示，称之为铁-应答元件(iron-response element，IRE)，因而把特异与之结合的阻遏物蛋白称为 IRE-结合蛋白(IRE-binding protein，IRE-BP)，但也有把它叫做铁调节因子(iron regulatory factor，IRF)的。当细胞内铁含量低时，IRE-BP 结合在 IRE 上，阻止了铁蛋白 mRNA 的翻译；当铁含量高时，则 IRE-BP 由 IRE 上解离下来，使铁蛋白 mRNA 得以翻译。那么铁的含量是怎样发挥这种调节作用的？最近已克隆了 IRE-BP 的全长 cDNA。分析其序列发现它与顺乌头酸酶有很高的同源性，而且此阻遏物本身也具有顺乌头酸酶活性。顺乌头酸酶含有个易变的[Fe-S]复合体，这提示 IRE-BP 是靠此复合体觉察细胞内铁的含量的。实验表明 IRE-BP 确实是个含[Fe-S]复合体的蛋白质。当在体外用铁盐处理它时可激活其酶活性，此激活过程是由于它从[3Fe-4 S]型转变为[4Fe-4 S]型的结果。而且类似的处理可使 IRE-BP 丧失其与 IRE 结合的能力。由此认为，当细胞中铁含量低时，IRE-BP 以[3Fe-4S]型存在，它结合在 IRE 上阻止翻译；当细胞内铁含量高时，它以[4Fe-4S]型存在，并从 IRE 上解离下来，使铁蛋白 mRNA 得以翻译。关于铁含量改变 IRE-BP 的作用机制还有其它解释，这里不再介绍。

已经研究了 IRE-BP 结合在 IRE 上阻止铁蛋白 mRNA 翻译的原因，有实验支持图 13-42 a 中的模式，即 IRE-BP 的结合阻止了核糖体 40 S 亚基与 mRNA 5′-末端的结合。虽然茎-环结构本身能够抑制翻译，但构成 IRE 的茎-环结构是不稳定的，这大概是它自己不能抑制翻译的原因。IRE-BP 的结合可以稳定 IRE 结构中的碱基配对，所以它的结合阻止了翻译。在铁蛋白 mRNAs 中，IRE 位于距 5′-帽子约 30 核苷酸处，这种与 5′-末端临近的位置是很重要的。因为当将此 IRE 移至距 5′-末端 67 个核苷酸的位置时，它就没有了阻遏作用。此实验也说明，在 40 S 核糖体亚基一旦结合在 mRNA 的 5′-末端上后，它的沿 mRNA 前进能够使其它蛋白质离去，并能使碱基配对结构解开。由于 40 S 核糖体亚基有清除其前进道路上的障碍的能力，因而一个 mRNA 结合蛋白(或者一个发夹结构)仅当与 5′-末端足够近，能够阻止核糖体 40 S 亚基的结合时，才能抑制翻译。虽然上述模式得到了一些实验的支持，但也提出了一些其它的模式，因而最后的结论还有待实验的进一步证明。

前已述及，在运铁蛋白受体(TfR)mRNA 的 3′非翻译区也有 IRE 结构，而且是用同一蛋白质——IRE-BP 与之结合，以稳定 TfR mRNA，增加 TfR 的合成。可见，在细胞内铁含量少时，IRE-BP 一方面结合在铁蛋白 mRNA 的 5′端 IRE 上，阻止其翻译，降低铁蛋白的合成；同时 IRE-BP 还结合在 TfR mRNA 的 3′端 IRE 上，增加 TfR 的合成。最近有实验提示，IRE-BP 还调控红细胞中 α-氨基-γ 酮戊酸合成酶 mRNA 的翻译，此酶是血红素生物合成途径中的第一个酶。当细胞内铁含量少时，IRE-BP 结合到此酶 mRNA 的 5′端 IRE 上，阻止其翻译，降低酶的合成。可见是由一个非凡的蛋白质同时调控着铁的含量恒定以及利用铁的一个关键代谢途径。

铁蛋白合成的调控是在真核中用阻遏物调控翻译的例子。虽然在原核中已经发现了很多用阻遏物调控翻译的例子，但在真核中除了铁蛋白以外还未发现其它确切的例子。而且在原核中已发现有用激活物调控翻译的例子，但在真核中，虽然从理论上说也应该有用激活物调控翻译的现象，但迄今还没有有说服力的证据。看来，对于真核基因表达的翻译调控还需要做很多的工作。

(二)酵母 GCN4 mRNA 翻译的调控

这是最近发现的一种真核的调控机制。啤酒酵母(*S. cerevisiae*) *GCN4* 基因编码一个转录因子，此转录因子活化 30～40 个基因的表达，它们的表达产物都是参与氨基酸的生物合成的。

GCN4 本身的表达则是在翻译水平上进行调控的，其 mRNA 是组成性存在的，但是仅当细胞缺乏氨基酸时它才被翻译产生出 GCN4 蛋白。GCN4 蛋白则启动一些基因的表达，使酵母进行氨基酸的合成，以满足其需要。如果酵母中有足够的氨基酸，则 *GCN*4 mRNA 不被翻译，酵母也就不进行氨基酸的合成了。

*GCN*4 mRNA 的翻译是怎样受到氨基酸的含量调控的？由于发现在 *GCN*4 mRNA 中，在其 5′-末端与 GCN4 的起始密码子之间有 4 个小的开放读框(open reading frame, ORF)而受到启发。已知在真核 mRNA 的翻译中，首先是核糖体的 40 S 亚基与 eIF-2-GTP · Met-tRNAi 形成复合体，此复合体结合在 mRNA 的 5′-帽子处，然后沿 mRNA 向 3′移动寻找起始密码子。当遇到离 5′-末端最近的起始密码子 AUG 时，则核糖体的 60 S 亚基结合上去，形成完整的核糖体，并由该 AUG 处开始翻译。可见真核 mRNA 的翻译总是由离 5′-末端最近的 AUG 开始的。实验表明，如果人为地在一个主要 ORF 的 AUG 上游插入一个起始密码子，则抑制了此 ORF 的正常翻译。然而，如果在上游 AUG 与主要 ORF 的起始密码子之间插入一个终止密码子，则能减轻上游 AUG 对主要 ORF 翻译的抑制作用。对此现象的最简单的解释是，在插入终止密码子后使主要 ORF 的上游形成一个 ORF，当 80 S 核糖体将上游 ORF 翻译完毕后，60 S 亚基解离出去，而核糖体的 40 S 亚基仍然结合在 mRNA 上，它再沿 mRNA 向 3′移动，寻找下一个 AUG，并能够由下一个 AUG 再启动翻译。就是说真核核糖体有再启动翻译(reinitiate translation)的能力。已发现，真核的再启动效率(与原核的不同)是随着上游 ORF 与下游 ORF 之间的距离加大而增高的。对于真核中这种基因间的距离越大越容易再启动翻译的合理解释是，当核糖体将第一个 ORF 翻译完，60 S 亚基离去后，40 S 亚基再开始寻找下一个 AUG 时，最初它大概没有识别 AUG 的能力，因为此时的 40 S 亚基没有 Met-tRNAi。两个 ORF 之间的距离加长，可为 40 S 亚基再获得 Met-tRNAi 提供了足够的时间，从而使它易于再启动翻译。此概念是说明酵母 *GCN*4 基因表达的调控机制的基础。

图 13-43 示出了酵母 *GCN*4 mRNA 翻译调控的概要。由于在 mRNA 的 5′-末端与 GCN4 的起始密码子之间有 4 个小 ORFs(图中仅示出了第 1 和第 4 两个)，因而 GCN4 的翻译必定有再启动机制的参与。对此体系的遗传学和生物化学的研究揭示出来了其调控途径的概要。在此途径中，细胞内 eIF-2-GTP · Met-tRNAi 的浓度调控着翻译的再启动，而细胞内 eIF-2 · Met-tRNAi 的浓度则是由氨基酸的含量所调控。如图 13-43 中所示，由核糖体 40 S 亚基、Met-tRNAi 和起始因子(包括 eIF-2)组成的复合体结合在 *GCN*4 mRNA 的 5′-末端，然后沿mRNA 寻找 AUG。当遇到第一个 ORF 的 AUG 时，核糖体的 60 S 亚基结合上去并开始翻译第一个 ORF。当第一个 ORF 翻译完后，60 S 亚基离去，40 S 亚基继续向下游寻找另一个 AUG。开始时由于 40 S 亚基中还不带有 Met-tRNAi，因而它没有能力识别 AUG，只有当它沿 mRNA 移动一定时间后，才能再结合上 Met-tRNAi，才有能力再启动翻译。图 13-43 a 表示在不缺乏氨基酸的情况下，有活性的 EIF-2 的含量较高。由于 EIF-2 的作用是将 Met-tRNAi 带到 40 S 亚基上去，因而能使翻译完上一个 ORF 的 40 S 亚基很快地结合上 Met-tRNAi，以获得再启动翻译的能力。因此，在不缺乏氨基酸的情况下 ORF4 被翻译。ORF4 翻译的终止位点以某种尚不清楚的因素排除了再启动翻译的进行，因此，在不缺乏氨基酸的条件下 GCN4(mRNA 中的第 5 个 ORF)不被翻译。而仅当核糖体能够越过 ORF4(不翻译 ORF4)时才能翻译 GCN4。图 13-43 b 所示是缺乏氨基酸的情况。当氨基酸缺乏时，由于未负荷的(uncharged)tRNA 的聚存引起了一个蛋白激酶的活化。此激酶使一部分 eIF-2 分子磷酸化而灭活(机制见后)，从而使 40 S 亚基再获得

Met-tRNAi 的时间加长，亦即在翻译完一个 ORF 后能够再启动下一个 ORF 的翻译所需的时间延长了。这就使得一些 40 S 亚基在到达 ORF4 时不能启动翻译，从而越过了 ORF4。但当这些 40 S 亚基继续前进至 GCN4 的起始位点时，它们有了足够的时间获得 Met-tRNAi，就能够启动 GCN4 的翻译了。这就是在缺乏氨基酸时能够表达 GCN4 的原因。

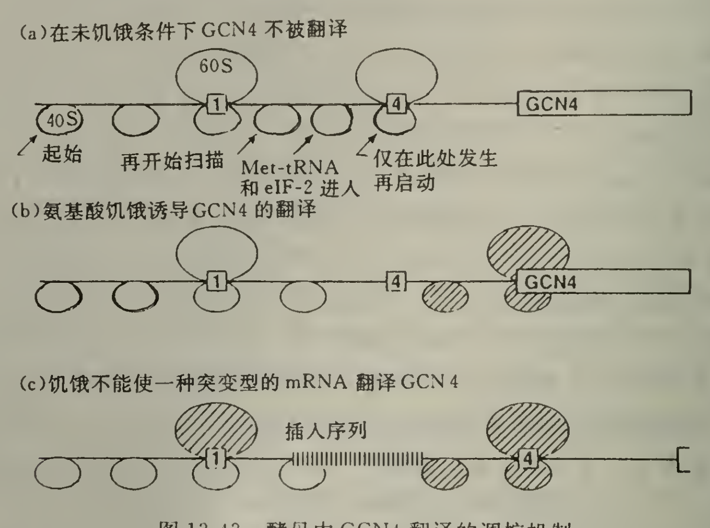


图 13-43　酵母中 GCN4 翻译的调控机制

●表示带有 Met-tRNAi 的 40 S 亚基，0表示不带有 Met-tRNAi 的 40 S 亚基。

图 13-43 所示是当人为地在 ORF1 和 ORF4 之间插入一段序列，使 ORF1 和 ORF4 之间的距离由原来的 200 个核苷酸变为 346 个核苷酸时，则虽然在缺乏氨基酸的条件下 GCN4 也不能翻译。对此实验结果的解释是，虽然由于氨基酸缺乏而使 40 S 亚基获得再启动能力的时间延长了，但由于 ORF1 和 ORF4 之间的距离加大了，因而当 40 S 亚基到达 ORF4 的 AUG 时仍能再启动翻译。ORF4 的翻译抑制了 GCN4 的翻译。因此，此实验结果支持上述关于 GCN4 翻译调控的模型。

已在植物和动物病毒以及脊椎动物的 mRNA 中发现有类似的情况，即在主要开放读框的前面有小的开放读框。然而对于这些 mRNA 中的小的开放读框的作用是什么目前还不清楚，更不用说其作用机制了。但已知远非所有的上游小的 ORF 的作用都是和 *GCN*4 mRNA 的一样。例如，酵母的另一个基因，*CPA*1 基因的 mRNA 也有个上游的小 ORF，但它的调控机制显然与 GCN4 的不同。*CPA*1 mRNA 的上游 ORF 编码一个 25-氨基酸的多肽，而不简单地只是像 *GCN*4 mRNA 那样是个负责调控翻译的上游结构。*CPA*1 的上游多肽阻遏 CPA1 翻译的具体机制尚不清楚。

引人注意的是，在脊椎动物中具有上游小 ORF 的 mRNA 并非是随意的。已知具有这种上游小 ORF 的 mRNA 的大多数是编码致癌蛋白(oncoprotein)，生长因子(growth factor)，转录因子，受体和其它关键性调控蛋白的 mRNA。目前尚无证据表明上述再启动的机制在这些 mRNA 的翻译中有重要的作用。这些 mRNA 的上游 ORF 的作用是什么？相信经过研究是会揭示出其奥秘的。

（三）**eIF-2 的磷酸化对翻译的普遍抑制**

酵母中 *GCN*4 基因表达的调控是近来才知道的利用 eIF-2 的磷酸化进行的翻译调控。然

而多年来即已知道，在哺乳动物细胞中通过 eIF-2 的可逆磷酸化调控翻译。例如，早就知道，当网织红细胞中血红素缺乏时，引起一种蛋白激酶的活化(血红素是此激酶的抑制剂)，此激酶使 eIF-2 的 α 亚基磷酸化而灭活，从而停止了所有蛋白质的合成[这就是在利用网织红细胞裂解液(reticulocyte lysate)进行体外翻译试验时必须向其中加入血红素的原因]。目前已知除了血红素缺乏外，还有许多因素可引起 eIF-2 的磷酸化，例如，不供给培养的细胞血清、热休克、氨基酸饥饿以及某些病毒感染等(图 13-44)，而且对其生物化学过程已经进行了仔细地研究。如图 13-44 所示，eIF-2 与 GTP 形成一个复合体，在溶液中它与 Met-tRNAi 结合在一起，然后 eIF-2 将 Met-tRNAi 带到核糖体 40 S 亚基上形成一个 43 S 的起始复合体。下一步是 43 S 复合体结合在 mRNA 的 5′-末端，沿 mRNA 向下游移动寻找起始密码子 AUG，在那里 eIF-5 结合上去，触发结合在 eIF-2 上的 GTP 水解。eIF-2 是个典型的 G 蛋白，GTP 水解后使其变为 GDP 型而发生构象的改变。GDP 型 eIF-2 与 40 S 亚基的亲和力比较低，因而 eIF-2-GDP 由 40 S 核糖体亚基上解离下来，而将 Met-tRNAi 留下。然后 60 S 亚基才能够结合上去，组装成有功能的 80 S 核糖体，即可进行翻译了。

由于 eIF-2 对 GDP 的亲和力比对 GTP 高 400 倍，所以需要一个辅助因子推动 GTP 和 GDP 的交换，以便由 eIF-2・GDP 再产生 eIF-2・GTP，此辅助因子叫做 eIF-2B，或者叫做鸟苷酸交换因子(guanine nucleotide exchange factor，GEF)。这些反应见图 13-44(b)。在蛋白质合成中，此 eIF-2 循环是个调控点，因为在多种不同的生理条件下见图 13-44(c)，eIF-2 的 α 亚基磷酸化了，这样阻止了 eIF-2・GTP 的再生成，从而也就阻止了所有 mRNA 的翻译。eIF-2 磷酸化后为什么阻止了 eIF-2・GTP 的再产生？这是因为细胞中 GEF 的含量很有限，而且它与 eIF-2(α p)的亲和力比未磷酸化的 eIF-2 高约 150 倍。因此，eIF-2 仅有约 25％磷酸化时，便能使所有的 GEF 与 eIF-2(α p)形成稳定的 eIF-2(α p)・GDP・GEF 复合体。这样不但结合在复合体中的 eIF-2(α p)不能再启动翻译，而且由于所有的 GEF 都束缚在此复合体中了，也就无法促使 eIF-2・GDP 再变为 eIF-2・GTP，因而停止了所有蛋白质的合成见图 13-44(c)。

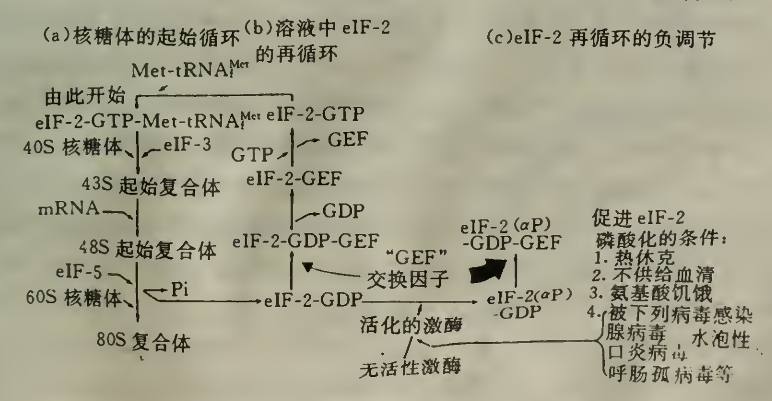


图 13-44　哺乳动物细胞中的 eIF-2 循环

(a)eIF-2・GDP 由 40 S 亚基释放出来，它是稳定的；(b)GEF 促使 GDP 和 GTP 交换；(c)eIF-2 磷酸化为 eIF-2(α p)，使 GEF 与之形成稳定的复合体，从而降低了整体的翻译能力。

已鉴定出两个催化 eIF-2 磷酸化的特异激酶。其中之一叫做 DA1，或者叫做 p68 激酶，它是由双链 RNA 活化的；另一个叫做 HR1，它的活化机制尚不清楚。但已知在网织红细胞中具有 HR1，它被血红素所抑制，因此，当缺乏血红素时 HR1 活化，使 eIF-2 磷酸化，从而封阻了网

织红细胞中的蛋白质合成。这两个激酶的 cDNA 已被克隆和测序。

关于催化已磷酸化的 eIF-2 脱去磷酸的特异磷酸酶目前知道的很少，然而可以肯定它必定参与此调控机制。另外已发现还有抑制 eIF-2 磷酸化的因子在起作用。已知通过 eIF-2/GEF 调控翻译的事实不仅限于哺乳动物，而是相当普遍地存在。因而对此调控系统的研究还有待深入。除了 eIF-2 外，是否还有通过其它起始因子以及延伸因子或核糖体蛋白质的磷酸化-脱磷酸化来调控翻译的？在这方面也已做了不少工作。虽然已经发现它们中有些蛋白因子确实发生磷酸化，但不能肯定这种磷酸化是否在调控中真的发挥作用，因而也是需要研究的问题。

在前面的叙述中，我们强调了通用起始因子的受到限制，例如 eIF-2 的磷酸化，对所有蛋白质合成的抑制作用。然而这种抑制作用的程度是否对所有蛋白质的合成都一样？会不会有些蛋白质的合成不受到抑制？即这种抑制作用不是对全部蛋白质的合成，而是有目标的？这是可能的。我们可以设想有 3 种可能的途径做到这一点。

(1)某一 mRNA 是通过特殊的机制进行翻译的，此特殊机制使得在活性通用起始因子的量不足时(不是绝对没有)，它的情况与其它大多数 mRNA 不同。例如前面讨论过的 *GCN*4 mRNA 就是如此。*GCN*4 mRNA 的翻译是靠再启动的，在 40 S 亚基再开始扫描至其到达下一个 AUG 之间是有固定时限的。在此期间，40 S 亚基必须结合上 eIF-2・GTP・Met-tRNAi 才能由下一个 AUG 处再启动翻译，如果结合不上，就只能越过它而不翻译。因此，在 eIF-2 部分失活时，使得所有蛋白质的合成速度都降低，但对 GCN4 的表达可以没有影响。

(2)激酶或磷酸酶的活化可发生在胞质溶胶(cytosol)的局部区域内，从而仅影响位于该区域内的 mRNA 的翻译，而不影响位于其它位置的 mRNA 的翻译。目前对于细胞溶胶的这种分区的了解还不充分，因而还不能具体说明此问题。然而已有实验证明，eIF-2 的磷酸化对有些 mRNA 翻译的抑制作用比其它 mRNA 大，而且造成这种差异的原因是由于其翻译被抑制的靶 mRNA 和引起磷酸化的诱导物(大概是双链 RNA)位于同一局部，而它们位于同一局部的原因尚不明了。

(3)在哺乳动物的 mRNA 中，其前导序列(leader sequence)形成二级和三级结构的倾向性是不同的，其结果使得它们提供 5′-末端的情况不同，因而在 43 S 起始复合体的含量下降时，有的 mRNA 能与之结合，有的则难于结合，从而使它们的翻译速度有所不同。已有证据表明，mRNA 的 5′非编码区的二级结构的程度与其所处位置都影响核糖体与 mRNA 的相互作用。然而遗憾的是，此学说难以用天然的 mRNA 进行验证，其原因是天然 mRNA 的 5′前导序列可被设想出多种不同的构象，而不是一种二级结构，因而其在细胞中的具体结构情况难以知道。

在 mRNA 的结构影响其翻译效率方面已经做了很多研究。例如，Kozak 分析了已发表的 620 个脊椎动物 mRNA(根据其 cDNA)的 5′前导序列中 GC 的含量，因为 GC 含量越高通常可以说明它能够形成二级结构的量越大。在这些 mRNA 中只有一小部分的前导序列所含 GC 的量＜50%，设想它们形成比较少的二级结构。具有这种前导序列的是编码珠蛋白、酪蛋白、清蛋白、α-胎蛋白(α-fetoprotein)和组蛋白等的 mRNA，所有这些都是大量产生的蛋白质，它们的 mRNA 是被有效地翻译的。与之相反，大多数脊椎动物 mRNA 的前导序列中含有较多的 GC(50%～70%)，提示它们形成较多的高级结构，因而在通用的翻译能力下降时，它们更少地被翻译。事实是，在除去细胞培养基中的血清、热休克或病毒感染时，大多数细胞的 mRNA 的翻译被抑制。极端的情况是有些 mRNA 的 5′非编码区中含有 70%～90%的 GC。典型的这种 mRNA 是编码致癌蛋白、生长因子、信号转导成分(signal transduction component)和转录因

子，以及许多持家蛋白质(housekeeping protein)。这些蛋白质的含量低，与它们的mRNA难以翻译的设想相符合。提高43 S起始复合体的含量大概也不足以促进这些GC含量非常高的mRNA的翻译。因为虽然43 S起始复合体一旦结合在mRNA上后，在它沿mRNA扫描时能够破坏沿途的二级结构而通过，但这种能力也不是无限的。它可能通不过上述这些GC含量非常高的前导序列中强有力的碱基配对结构。可见，要想提高生长因子等的翻译效率需要加入RNA解旋酶(RNA helicase)，此酶应能使43 S起始复合体易于进行扫描。然而迄今尚未找到这样的酶。另一个办法就是改变它们的前导序列。

目前已积累了一些关于mRNA的结构影响其翻译效率的资料。除了上述5′非编码区的结构对翻译有影响外，其3′非编码区(例如有的结构影响mRNA的稳定性，见前)以及编码区(例如使用同一氨基酸的不同密码子)的结构都对mRNA的翻译效率有影响，而且它们的影响是多种多样的。查明这些现象不仅具有理论意义，而且还有实用价值。因为当我们用真核细胞体系表达某一基因以期获得其产物时，总是希望其表达效率越高越好。很清楚，要想得到高效表达，不但需要注意转录中的问题，还必须考虑如何提高其mRNA的翻译效率的问题。

第二节　细胞周期和生长的调控

一、细胞周期

分裂活动是从上一次细胞分裂后开始发生的一系列事件的顶点，两次细胞分裂之间的时期称为体细胞周期(cell cycle)。从上次有丝分裂的终止到下次分裂开始之间的时期称为间期(interphase)。实际分裂的期间，即相当于可以看到有丝分裂的期间称为M期(mitosis phase)。一个细胞周期也可停止于减数分裂，产生具有单倍体染色体成分的配子。在有丝分裂和减数分裂的调控中有共同的情况。

为了分裂，一个真核体细胞必须使其物质加倍，然后将其组成成分等量地分配给两个子细胞，即产生两个相同的子细胞。体积的加倍是个连续过程，这是编码形成特定细胞表型的蛋白质的基因转录和翻译的结果，所以RNA和蛋白质的合成是不断进行的。与之相反，基因组的复制则仅仅发生在特定的DNA合成期，称之为S期，S是表示合成期(synthetic period)之意。所以DNA的合成是间断的。因此，根据DNA合成的时间又把间期分成为几个期，如图14-45所示。在有丝分裂完成后细胞进入第一个间隔期(gap phase)，称为G1期。此时RNA和蛋白质在合成，但不发生DNA复制。DNA复制的开始意味着G1期过渡到S期。在S期

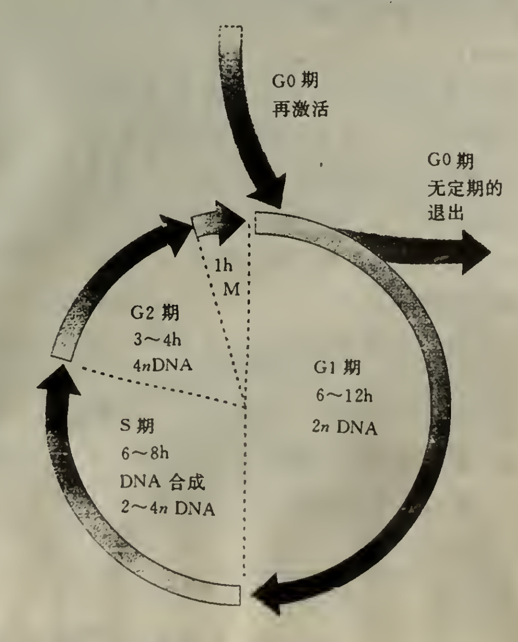


图 13-45　间期分为G1、S和G2期。一个细胞周期和下一个细胞周期之间用有丝分裂划界细胞可由周期退出为G0，也可由G0再进入周期。

DNA 持续地合成，直到完全复制出来为止。由S期终了到有丝分裂开始前是第二个间隔期，称为G2期，在此期间细胞含有两套二倍体的染色体。

由此可见，在整个间期细胞的外表虽然看不出什么变化，但其RNA和蛋白质的量在不断增加，DNA则在S期加倍（此时细胞核的体积显著增大），但染色质的状态没有看得出的变化，最后各种物质都加倍，就为有丝分裂做好了准备。需要指出的是有些细胞周期与上述体细胞的有所不同。例如，在胚胎形成期可发生特异的细胞分裂，分裂出的细胞是彼此不同的；又如卵的生成，它要聚存足够的物质以维持其早期分裂。

在有丝分裂时分给每个子细胞一套二倍体染色体。在此期间细胞结构发生深刻的改变：核包膜裂解了；染色质发生凝聚，仅在此期间各个染色体变为可以看到的了；细胞骨架发生深刻的再组建，细胞在纺锤体上再建等。在有丝分裂期间所有的合成过程实际上都暂停了。

一个正在进行周期活动的动物体细胞，其周期活动是每18～24 h重复一次。图13-45中表明G1期一般最长，它由快速生长的动物细胞的约6 h到较慢生长细胞的约12 h之间变动。细胞周期的长短不同主要是由于G1期长短不同之故。S期的长短由复制全部基因组所需的时间而定，典型的为6～8 h。G2期通常是间期中最短的，它大概是有丝分裂的准备阶段。M期通常不到1 h。

二、两个关键的决定位点

细胞周期是怎样调控的？已发现在细胞周期中有两个位点决定着周期是否继续进行。第一个位点是在G1期，此位点决定着染色体是否复制，只有条件成熟时才开始DNA的合成，即进入S期。所谓条件成熟大概有评估培养基的状况以及细胞物质的含量是否足够等等。此决定位点在酵母中研究得最清楚，把它称为起点（start），在动物细胞中称之为限制点（restriction point）。第二个位点在G2期，此位点决定着是否进行有丝分裂，即进入M期。检验是否可以进入M期的条件有：细胞物质的含量是否已增加到足够分裂；复制是否已经完成和DNA有没有损伤等。对于培养的动物细胞来说，已知G1调控是主要的决定位点，而G2/M调控是辅助的。细胞在其周期中在G1期花了最长的时间，而且调节G1期的长短反映着生长的条件，一旦细胞经过了G1期，则除非发生偶然情况，它将完成S期，经过G2期，并进行分裂。培养的细胞不会停止在G2期。现在看来，大多数二倍体细胞大概都是在G1期调控，培养的和体内的细胞都是如此。

有些细胞表型是永不分裂的。这些细胞被认为是由细胞周期中撤出而成为另一种状态，此状态类似于G1，但又不同于G1，因为它不能进入S期，此种状态称为G0。有些类型的细胞可被刺激由G0再进入细胞周期。由细胞周期的撤出或再进入，实际上是发生在G1期的早期阶段。有些类型的细胞停止在G2期。在二倍体细胞中，这种类型的细胞通常是要进行分裂的，例如，在昆虫胚胎发生的某些阶段，核分裂了，但停止在四倍体状态。单倍体细胞更常常停止在G2期，这提供了保护DNA不受损伤的作用，因为此时以双拷贝基因组代替了G1期的单拷贝基因组。有些酵母可用G1期或G2期作为主要调控位点，这取决于营养状况。

三、在S期和M期有起调节作用的物质

在早期的实验中已经发现在细胞周期的不同阶段有不同的调控物质，这是将处于细胞周期不同阶段的细胞融合在一起发现的，如图13-46所示。融合后的杂交细胞可在共同的胞浆中

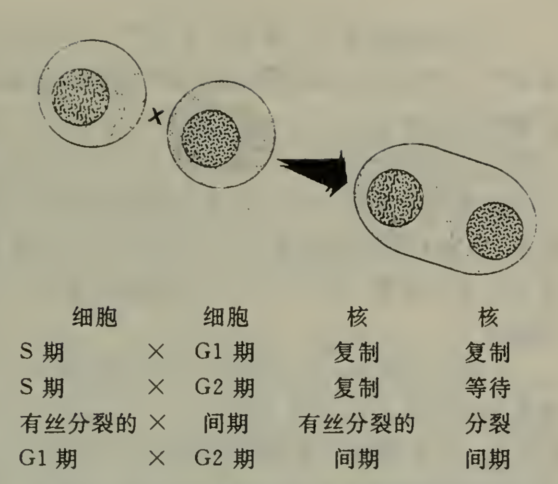


图 13-46 细胞融合后的杂交细胞的状态取决于被融合的细胞的状态

含有两个或更多个核。从图中可以看出：(1)当一个S期细胞与一个G1期细胞融合时，产生的杂交细胞中的两个核都复制DNA。这提示在S期细胞的胞浆中有一个DNA复制的激活物，叫做S期激活物(S phase activator)；(2)当一个S期细胞与一个G2期细胞融合时，S期细胞核继续复制，但G2期的核不复制。这提示已经复制了的DNA变为对S期激活物不起反应了，这保证了每个DNA分子仅复制一次。然而在这种融合中看到了有丝分裂在时间上发生了改变，即S期核比在它原来的胞浆中进入M期要快些，而G2期细胞则要等到S期核复制完毕后才进入有丝分裂。这可能意味着在S期细胞中有某种调节物抑制着有丝分裂的开始，此调节物大概就是S期激活物本身；(3)当一个有丝分裂的细胞与一个处于间期任何阶段的细胞融合后，它引起间期细胞核进入一种假有丝分裂(pseudomitosis)，其特点是染色体的超前凝聚(premature condensation)，提示在正处于分裂中的细胞内有个M期激活物；(4)当G1期和G2期细胞融合后，杂交细胞中的任何核都不被诱导进行复制或有丝分裂，可见在G1期和G2期细胞中不存在S期激活物和M期激活物，说明这些激活物的活性是暂时存在的，S期和M期一过，它们便分别被灭活。

下面我们将讨论S期激活物和M期激活物的本质及其调控机制。虽然对于这些调控机制目前还未完全清楚，然而可以看出，通过调控保证了在细胞周期中必须是在上一个阶段完成后才开始进入下一阶段。另外，关于真核细胞周期的调控主要是通过爪蟾的卵母细胞的生化分析和酵母细胞周期的遗传分析进行的，然而已知调节真核细胞周期的主要调控成分是高度保守的，所以用这两种生物揭示出来的调控机制有着足够的代表性。

四、由爪蟾卵中制备出促成熟因子

爪蟾(*X. laevis*)卵的早期发育提供了一个分析细胞如何进入有丝分裂的非常有用的体系。爪蟾的卵母细胞在其未成熟型时被抑制在其第一次减数分裂周期中，正好在染色体凝聚的开始阶段，相当于体细胞的G2期。当激素触发它进入第二次减数分裂时发生排卵。当卵排出后它被阻止在接近第二次减数分裂的末尾，其状态相当于体细胞的M期。

由于卵的体积比较大(爪蟾的卵直径约为1 mm)，特别是由于休止的卵母细胞和休止的卵都易于获得，因而提供了一个很方便的研究体系。实验表明，将休止的卵的胞浆抽提物注射给休止的卵母细胞时，促进了卵母细胞的有丝分裂，可见在卵细胞中有一种活性物质，它能使未成熟的卵母细胞由G2期进入到M期，故将此物质称为促成熟因子(maturation promoting factor MPF)。后来证明MPF有普遍的引起体细胞进入M期的作用，所以就把它叫做M期促进因子(M phase promoting factor，代号仍是MPF)。已证明MPF的活性随着有丝分裂的过程发生变动，即在发生有丝分裂时MPF的活性升高，而在有丝分裂已经完成和S期时它的活性消失了。

已由爪蟾的卵中分离出来了MPF。它是由两个亚基组成的二聚体，具有蛋白激酶活性，能使多种不同的蛋白质磷酸化。MPF 的两个亚基之一的分子量为 34 kD，故称为 p34，它是起蛋白激酶作用的亚基。另一个亚基的分子量为 45 kD，故称为 p45，它是一种周期素(cyclin)。p34 的蛋白激酶活性需要 p45，它必须与 p45 结合后才有激酶活性。周期素的种类很多，将这类物质叫做周期素是因为它们在整个细胞周期中不断增多，而在有丝分裂过程中突然被蛋白酶降解掉。有丝分裂的周期素可分为两大类，即周期素 A 和周期素 B。目前对这些周期素的功能各是什么还不清楚，但已知 MPF 中的周期素是 B 类周期素。MPF 的更为合适的名称应该是 M 期激酶。

在由 G2 期进入 M 期时活化 M 期激酶以及在 M 期中灭活它是细胞周期中的关键。M 期激酶的活化和灭活循环见图 13-47。已知在大多数细胞的整个细胞周期中 p34 的含量是保持恒定的。在 S 期新合成的周期素与 p34 结合起来，形成 M 期激酶，但这时它是没有活性的。然后被不同的酶先后在 p34 的苏 14、酪 15 和苏 161 处磷酸化，此时它仍没有活性，但易于被触发。关键的一步是由一个磷酸酶将 p34 中苏 14 和酪 15 上的磷酸基脱掉，这样 M 期激酶就有了活性，从而启动了有丝分裂。此磷酸酶是 cdc25 蛋白(p34 也叫 cdc2 蛋白，这些蛋白质的名称来源见后)，它是在 DNA 合成完成后才起作用的。

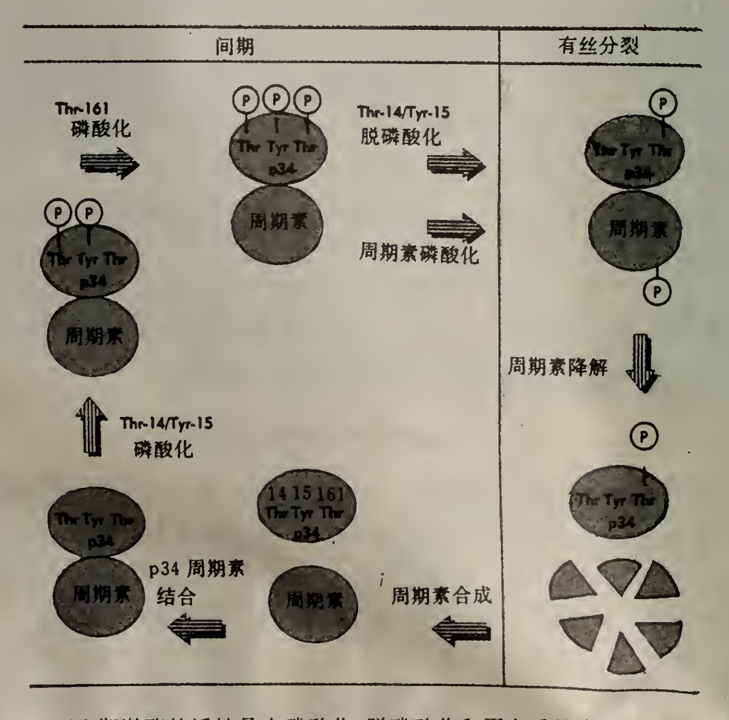


图 13-47　M 期激酶的活性是由磷酸化、脱磷酸化和蛋白质的降解作用调控的

磷酸化的次序是苏-14，酪-15，苏-167。各种事件发生的时间尚未确定。

M 期激酶的灭活是细胞离开有丝分裂所必需的，是完成有丝分裂的先决条件。M 期激酶是怎样灭活的？是通过其周期素亚基的降解实现的。周期素亚基降解的机制尚未完全明了，但已知是 M 期激酶自己启动的，它使周期素产生一个新的氨基末端，从而使之与泛素(ubiquitin)结合，而泛素的结合是标志出被它结合的蛋白质要被降解的。可见 M 期激酶的活

性是自我限制(self-limiting)的,这样就自动地在有丝分裂持续时间的长短上安置了一个界限。

五、蛋白质的磷酸化和脱磷酸化调控着细胞周期

由上述可见,M 期激酶的活化触发了有丝分裂的开始。在有丝分裂时细胞的结构发生深刻地变化,其中主要有:①染色质凝聚为可识别的染色体;②核纤层(lamina)的溶解和核包膜(envelope)的裂解;③微管(microtubules)的解聚和再构建成纺锤体;④为了胞质分裂(cytokinesis)而发生肌动蛋白丝(actin filaments)的再组织。有丝分裂后,M 期激酶灭活,这些细胞结构又恢复到间期的状态。可见 M 期激酶调控的都是可逆的过程。

对于 M 期激酶调控这些过程的细节虽然还不清楚,然而已知是 M 期激酶活化后,使得一些蛋白质发生了磷酸化,从而诱发出上述在有丝分裂时发生的细胞结构的各种变化。在有丝分裂后,M 期激酶灭活了,被 M 期激酶磷酸化了的各种蛋白质又在磷酸酶的作用下脱去磷酸基而恢复原状,从而使得细胞的结构恢复为间期的状态。有证据表明,在由 G1 期进入 S 期的调控机制也是通过某些蛋白质的磷酸化和脱磷酸化进行的,所以说是蛋白质的磷酸化和脱磷酸化调控着细胞周期。

结合体外和体内的实验表明,M 期激酶使下述蛋白质发生磷酸化可能是引起有丝分裂时细胞结构发生改变的原因:①H1 组蛋白的磷酸化可能是引起 M 期染色体凝聚的原因;②核纤层的组成单位是核纤层蛋白(lamin)。M 期激酶可使核纤层蛋白磷酸化,从而使核纤层解聚为核纤层蛋白,引起核的完整性的丧失和核包膜的破裂。关于引起细胞其它结构发生改变的蛋白质尚在鉴定之中,催化这些蛋白质于磷酸化之后再脱去磷酸基的究竟是些什么磷酸酶也是需要解决的问题。

六、酵母的遗传分析揭示出多种与调控细胞周期有关的基因

为了确切说明细胞周期调控的复杂机制,需要鉴定出那些直接参与调控细胞周期的基因。其法是找到一些细胞的突变体,由于突变而把这些细胞封阻在细胞周期的某一特定阶段,从而得知此突变的基因是与该特定阶段的调控有关的基因。由于这种突变对于动物细胞来说是致死的,所以无法用以进行研究。然而用酵母则可以进行,所以迄今已用酵母做了大量的工作。由于已知细胞周期的调控机制在进化中是非常保守的,所以也就揭示出来了许多有关真核细胞周期调控的原理。

酵母的有丝分裂循环有两种方式,如图 13-48 所示。*裂殖酵母*(fission yeast, *S. pombe*)的细胞周期和动物细胞一样,也分为相同的期。细胞先纵向生长,然后分裂。其通过细胞周期的进度大致可用细胞的长度(由 7～8 μm 到加倍为 14～16 μm)来评估。面包酵母(baker's yeast, *S. cerevisiae*)的周期是不一般的。细胞几乎是由 S 期直接进入分裂,G2 期实际上很短或没有(图中标出了 G2 期是为了标出向 M 期过渡的时相)。并且细胞不是相等的分裂,而是子细胞以由母细胞芽出的方式生长,最后以一个较小的细胞释出而获得独立。

酵母细胞可以单倍体或二倍体形式存在。它们有两种型式的生命周期。单倍体细胞可经有丝分裂而加倍。一个单倍体细胞可以是交配型 a 或 α。相反型的单倍体可进行性交配途径,在此途径中它们接合(conjugate)起来形成二倍体细胞,二倍体细胞再通过减数分裂形成单倍体细胞。

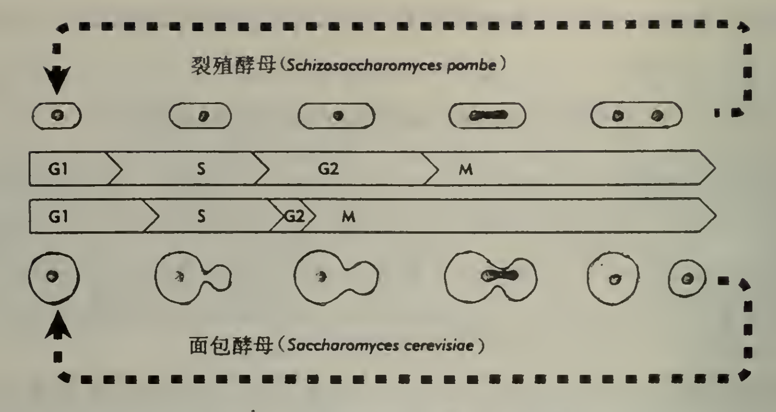


图 13-48　两种不同的酵母细胞周期

单倍体细胞的行为规定了细胞周期中的一个关键之点，即在其 G1 的早期阶段决定了究竟是进行分裂周期还是进行交配。这种决定是受环境影响的，例如，必须具有相反交配型细胞时才能发生接合作用。事实是由一种交配型的细胞分泌出一种交配因子(mating factor)，此因子是个多肽激素，它引起另一种交配型细胞停止其周期。但交配因子仅能在 G1 早期使细胞转向交配途径。在 G1 晚期细胞变成为进入分裂周期，不能再被刺激进入交配途径。G1 期中的这个决定点称为起点。

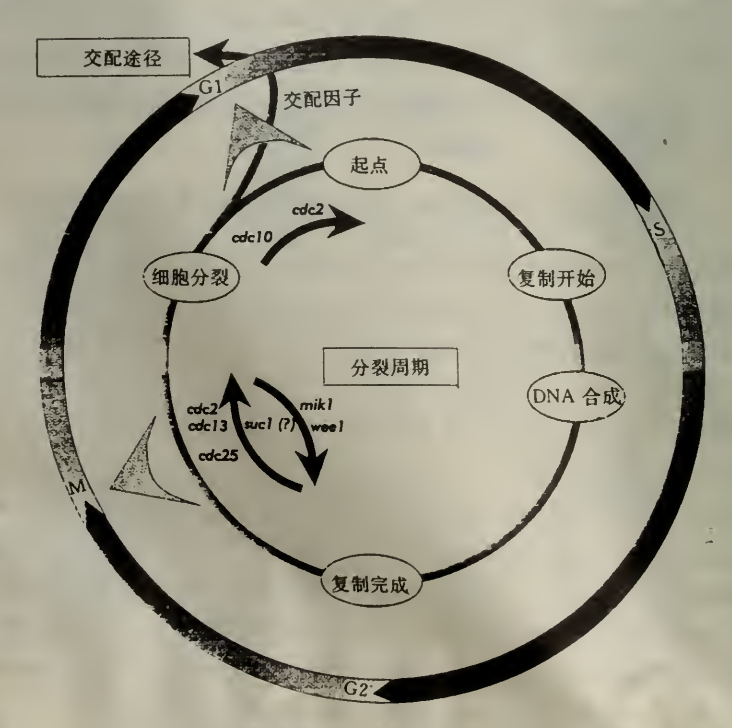


图 13-49　*S. pombe* 的细胞周期需要 *cdc*2 基因以通过特定的阶段，但可被反映细胞大小的基因(*weel*)所阻止，细胞可在 G1 的早期转变为交配途径

在两种酵母中都进行了广泛的细胞分裂周期(cell division cycle，cdc)的突变体的筛选工作，从而揭示出许多与调控细胞周期有关的基因，简称为 *cdc* 基因。图 13-49 概括了 *S. pombe*

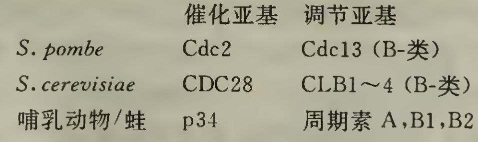
周期中所涉及的事件和基因。*S. pombe* 的各种 Cdc 突变体或在 G2 和 M 期的边界上，或在 G1 中的起点处封阻细胞的周期。

已确定 Cdc2 蛋白(*cdc*2 基因的产物)是个关键的调节物，因为 *cdc*2 的突变体可被封阻在起点之前或 M 期之前，可见细胞周期的两个调控位点都是受 Cdc2 蛋白的调控。前已述及 Cdc2 蛋白是爪蟾 p34 的同源物，它和 p34 一样都是蛋白激酶，并通过它们的蛋白激酶活性引发细胞周期的向前进行。然而它们需要与其它蛋白质相互作用后才具有激酶活性。这些蛋白质都是 *cdc* 基因的产物。已知为了通过起点，除了 *cdc*2 外还需要 *cdc*10，此基因编码一个转录因子的亚基，此转录因子活化 S 期所需的基因。而为了由 G2 期进入 M 期，则除了 *cdc*2 基因外，还需要 *cdc*13 和 *cdc*25 基因，以及其它基因。

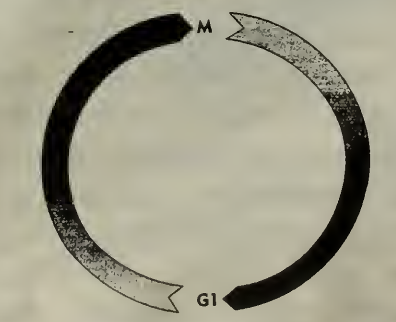
关于 Cdc2 蛋白在调控由 G2 期进入 M 期的作用机制知道得比较多。它与 *cdc*13 的产物(一种 B 类周期素)结合成为 M 期激酶。与图 13-47 中所描述的哺乳动物的 p34 的情况类似，Cdc2 蛋白的酪 15 必须脱去磷酸基，而苏 161 则需要磷酸化，但没有与哺乳动物 p34 相似的苏 14 的修饰作用。关键的一步是 *cdc*25 的产物即 Cdc25 蛋白的作用，这是个磷酸酶，特异地脱去 Cdc2 蛋白酪 15 的磷酸基，使 M 期激酶活化，从而引起有丝分裂。在有丝分裂时 Cdc25 蛋白的量增加了。看来 Cdc25 蛋白的量逐渐增加，当它的量超过一个阈值时才使 M 期激酶活化并引起有丝分裂是重要的。Cdc25 蛋白的活性似乎起着一种检查作用，它保证了只有当 S 期完成后才启动 M 期。另一个能与 Cdc2 蛋白结合的蛋白质是基因 *suc*1 的产物 p13，然而它的功能迄今尚不明了。

在正常情况下，细胞的分裂周期是与细胞的大小相关的。在不良的生长条件下细胞的增长较慢，G1 变得更长。这是个保护作用，用以避免在细胞的物质含量达到足以支持两个子细胞之前进行危险地分裂。已经鉴定出检验细胞大小的物质之一是 *weel* 基因的产物，此产物是个激酶，它能使 Cdc2 蛋白的酪 15 磷酸化，从而抑制 Cdc2 蛋白的活性。*mik*1 的产物起着同样的作用。*wee*1 和 *mik*1 都缺失是致死的。由此可见，*wee*1 和 *mik*1 基因的产物是检查细胞大小的，它们保证了细胞在不够大之前不进行分裂。

由上述可见，*wee*1 的产物和 *cdc*25 的产物起着相互拮抗的作用。*wee*1 产物的激酶活性作用于 Cdc2 蛋白的酪 15，抑制其功能，而 Cdc25 蛋白的磷酸酶活性作用于同一位点以活化 Cdc2 蛋白。Cdc25 过度表达的突变体与缺乏 *weel* 的突变体表型相同。这些工作所建立起来的基本原理是：关键的调节物是 Cdc2 蛋白，它的活性受到激酶和磷酸酶的调控，而这些激酶和磷酸酶的活性又受到其它信号(例如细胞的大小是否足够，DNA 的合成是否完毕等)的调控。所有这些信号都最后通过 Cdc2 蛋白的作用决定是否前进通过周期。



	催化亚基	调节亚基
S. pombe	Cdc2	Cdc13 (B-类)
S. cerevisiae	CDC28	CLB1～4 (B-类)
哺乳动物/蛙	p34	周期素 A,B1,B2



	催化亚基	调节亚基
S. pombe	Cdc2	Puc1
S. cerevisiae	CDC28	CLN1～3
哺乳动物/蛙	CDK2(p33)，CDK4	周期素 A,C D1,D2,D3,E

图 13-50 用于构建 M 期激酶和 G1 期所用激酶的可能成分

关于调控起点的知识比调控有丝分裂的知识要少些，这些知识首先是分析酵母 *S. cerevisiae* 的 *cdc* 突变体获得的，然后用以鉴定其它生物。已知在 *S. cerevisiae* 中调控通过起点和进入 M 期两个阶段的关键物质都是基因 *cdc*28 的产物（是 *S. pombe* 中 Cdc2 的同系物）。和 Cdc2 蛋白一样，Cdc28 蛋白也是蛋白激酶的催化亚基，需要与周期素结合起来才有作用。但在调控起点所用的周期素与调控进入 M 期所用的周期素是不同的，因而有时把它们分别称为 G1 周期素和 G2 周期素。图 13-50 中列举了两种酵母和哺乳动物中的调节性激酶的可能成分。由图中可见，真核生物调控细胞周期的一致策略是激酶由一个催化亚基和一个调节伙伴所组成。在有丝分裂时催化亚基由单个基因的产物（*cdc*2 和 *cdc*28 产生的蛋白质 p34）所提供，而调节伙伴则通常不是一种，而是由 B 型或者 A 型的周期素所提供。在调控由 G1 期进入 S 期时，酵母所用的催化亚基与 M 期激酶的相同，而在动物细胞中则是一个与 p34 有关的蛋白质，即 CDK2 或 CDK4。调节伙伴则是一大类 G1 周期素中的一种。这类周期素在序列上与有丝分裂时的周期素关系很小。当前关于周期素的定义已和原来所说的 A 型或 B 型周期素的定义有所不同，除了 A 型和 B 型外又发现了多种周期素，它们的性质也不都是像原来所描述的在细胞周期中逐渐积累，在有丝分裂过程中突然被蛋白酶降解。现在把与原来类型的周期素在序列上有一定相关性，并且能够与一个 p34 或 CDK2 以及相关的催化亚基结合起来形成一个激酶的蛋白质，都叫做周期素。

第三节　发育的基因调节

发育是由单个受精卵开始的，但由它产生的细胞中却有着不同的发育结局。早期发育需要了解的问题是，在分子水平上这种不对称是怎样引入的，即开始时的一个细胞，仅仅经过几次分裂，是怎样使得这些子细胞的性质彼此不同的？不同生物引入不对称的方法不同，当前仅对果蝇研究得比较清楚。已知果蝇的卵子本身是不对称的，这是由母方基因引起的，即在卵子发生期间，母体细胞中的一些基因的表达产物（包括 mRNA 和蛋白质）进入卵子，这些产物在卵子中做不均匀地分布，从而使卵子成为不对称。同时在受精后的初期仅发生核分裂，产生的许多核都处于同一胞浆中（称为合胞体）。随后这些核分布在合胞体的固定的不同部位处。由于母方基因的产物在合胞体胞浆中的分布是不对称的，因而处于不同位置的核的环境不同，而且母方基因的表达产物对基因的表达有调控作用，因而使处于不同位置的核（以及随后由它们形成的细胞）的基因表达的情况有所不同，这就使得细胞分化了。经研究，受精后开始表达的基因称为分节基因，它们的种类很多。其中有些是在母方基因产物的影响下分别在不同的细胞中局部表达，它们的表达又调控下一类分节基因的表达。分节基因共分为 3 类，在它们的作用下使胚胎分成若干个节。它们并且控制着下一群基因的表达，这一群基因称为同源异型基因。在不同节中表达不同的同源异型基因，同源异型基因的产物都是转录因子，它们在不同节中分别启动不同的一套基因进行表达，其结果使不同的节产生不同的表型。个体发育的调控过程很复杂，但上面的过于简单的描述说明了一个基本的原理，即个体发育是在一套基因控制下一套基因，下一套基因又控制再下一套基因的基因级联的方式下进行的，这个原理是所有生物都共同的。下面将根据上述过程主要讨论果蝇发育的已知的有关细节，最后简单说明其它生物也有与果蝇发育相似之处。

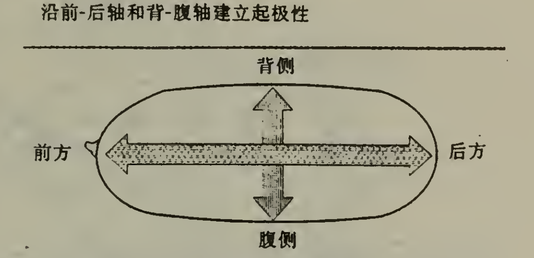
一、必须把一个梯度转变为分隔的分区

在图13-51中按发育的3个阶段，卵、幼虫和成虫表明了果蝇发育的基本问题。在发育开始时，沿着卵的前-后(anterior-posterior)和背-腹(dorso-ventral)两个轴(axis)建立起梯度(gradient)。卵的前端将变为成虫的头，后端将变为尾。背侧在上，腹侧在下。梯度是由分子(蛋白质或RNAs)在胞浆中差别分布形成的。担负前-后发育的梯度在受精后立即形成，而背-腹梯度的建立则稍晚一些。概括但过于简单的说法是：前-后体系调控沿幼虫体的位置信息，而背-腹体系则调控组织的分化，即规定不同的胚胎组织，其中包括中胚层(mesoderm)，神经外胚层(neuroectoderm)和背部外胚层(dorsal ectoderm)。

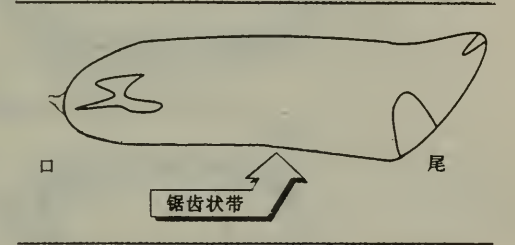
昆虫的发育分 2 个阶段，首先是由卵发育成幼虫，然后由幼虫变态(metamorphoses)为蝇。而幼虫(胚胎)和蝇的结构是十分不同的。随着幼虫的发育，它形成一些幼虫所特有的部位，这些部位不形成成虫的组织，而另些部位则变态为成虫的结构。与之相反，哺乳动物的胚胎则发育出与成年相同的身体各部位。尽管发育过程有这样大的不同，但控制昆虫发育和脊椎动物发育的基本原理看来是相同的，已发现了果蝇的调节因子和哺乳动物的调节因子之间的相关性。

分隔的区域是在胚胎中确定的，它们相当于成虫身体的各部位。在图 13-51 中根据幼虫的表面结构指出了这些分隔区。在幼虫的表皮(cuticle)上有特定花样的锯齿状突起(小茸毛)的带。其特点是：①沿前-后轴锯齿状突起(denticles)形成分隔的带，每个带相当于成年果蝇的一个节。事实是 11 条锯齿状突起的带相当于成虫的 11 个节；②沿背-腹轴的情况是，腹部的锯齿状突起粗，而背部的则细得多。虽然表皮仅仅代表体表，但它的结构是胚胎两个轴上整个机构的特征。因此，根据沿两个轴的锯齿状突起花样的畸变已做了很多在果蝇发育中突变体表型的分析。

卵
沿前-后轴和背-腹轴建立起极性



幼虫
口在前方，尾在后方。锯齿状带由腹侧扩出，并沿 前-后轴分出节



成虫
具有 3 个胸节和 8 个腹节的分节结构

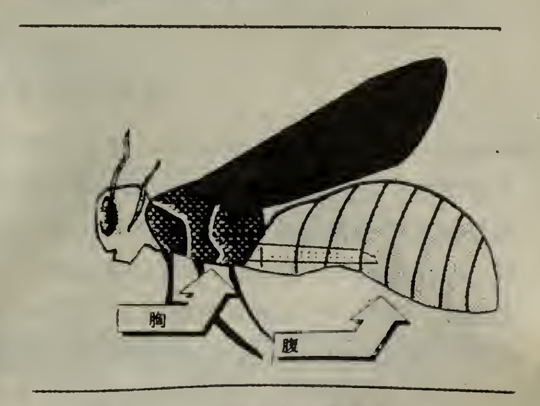


图 13-51　果蝇发育的基本情况

根据卵的梯度和成虫的节之间在形式上的差异提出了一些基本的问题。卵中的梯度是怎样建立起来的？一个连续的梯度是怎样转变成为决定每个细胞类型的不连续的差异的？由单一的梯度怎样能够发展成为数目很多的分开的分区？

梯度的本质以及它们影响位于整个胚胎中的多种细胞型发育的能力，是依靠果蝇发育的某些特性的。图 13-52 概括了其早期阶段。在受精的卵中有两个亲本的核，并且在其后端有个

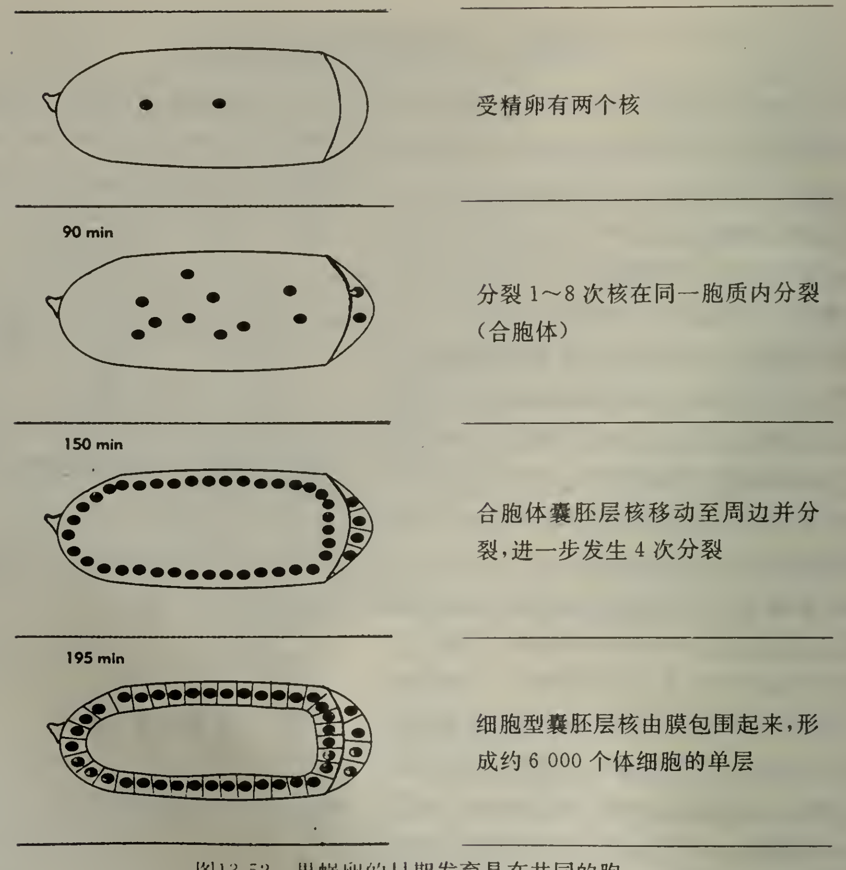


图13-52　果蝇卵的早期发育是在共同的胞浆中进行的，直至细胞型囊胚层阶段为止

叫做极质(polar plasm)的区域。在头 9 次分裂时核是在共同的胞浆中分裂的。在此胞浆中物质是能够扩散的(虽然可能有细胞骨架结构的约束)。在第 7 次分裂时有些核移动至极质中，在那里这些核成为生殖细胞的前体。在第 9 次分裂之后，核移动并分裂以便在卵的表面形成一层。然后它们再分裂 4 次，分裂后膜包围了它们而形成体细胞。直到细胞形成之前所有的核都是在同一胞浆中。在细胞型囊胚层(cellular blastoderm)阶段，分开的分区开始变为明显了，并且在此时卵的特定区域被确定为变成成虫的特定型结构(这种确定是逐渐的和顺序变化的)。然而在此过程开始时，当核移动至表面以形成囊胚层的单层时，每个核到达什么位置是完全随机的。由此可见，是核到达的位置决定了它们的后代细胞将变成什么型。

二、在胚胎发生的早期由母体基因的产物建立梯度

最初的不对称是在卵子发生(oogenesis)时加给果蝇的卵母细胞(oocyte)的。图 13-53 表明了果蝇卵巢中一个滤泡(follicle)的结构，这是一次减数分裂和随后 4 次有丝分裂以产生 16 个单倍体细胞的结果。其中一个细胞变成了卵母细胞，其余 15 个变成了滋卵细胞(nurse cell)。卵母细胞和与之相邻的滋卵细胞之间用胞质桥(cytoplasmic bridge)或者叫做环管的(ring canal)相连通。细胞的胞浆物质，包括蛋白质和 RNA 可进入卵母细胞，这些物质占卵的相当大的一部分体积。胞质桥位于卵母细胞的一端，此端将成为卵的前端。

在母蝇体内表达的基因对早期发育很重要，这些基因叫做母体基因(maternal gene)，可用雌性不育(female sterile)突变进行鉴定。由于这些基因不影响母蝇自己，但对产生后代是必需的，因而这种突变的雌蝇所产的卵不能发育为成虫，它们在发育期间死亡。

母体基因的共同特点是在受精前表达，尽管它们的表达产物有的是在表达之时起作用，有的则贮备起来在后期使用。按其表达的地点可将母体基因分为两类：①在母体细胞(例如滤泡细胞)中表达并影响卵发育的基因称为母体体细胞基因(maternal somatic gene)；②在种系细胞(滋卵细胞或卵母细胞)中表达的叫做母体种系基因(maternal germline gene)。

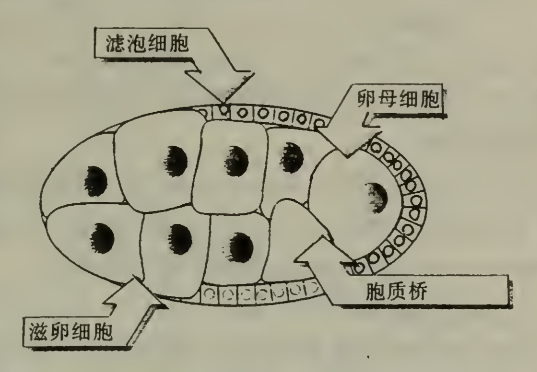


图 13-53　果蝇的一个滤泡，外面有滤泡细胞，它们包围着细胞和与之紧密接触的卵母细胞

滋卵细胞通过胞质桥彼此联结，并和卵母细胞的前端相结。滤泡细胞为体细胞。

用母体基因突变的方法鉴定出来了 4 组与胚胎的特定区域发育有关的基因。用常规遗传学测定和生物化学分析得知，每组基因都可按其活动次序组成一个途径，这些途径的成分列于图 13-54。有时把途径中活动较早的基因叫做上游基因，活动较晚的叫做下游基因。从图中可以看出，4 个途径运转的共同原理是，每个途径都是由局部事件开始，这些事件有的发生在卵内，有的发生在卵外，其结果是使卵内的一个信号局部化。此信号是个蛋白质，它以不对称的方式分布于卵内，称为形态发生素

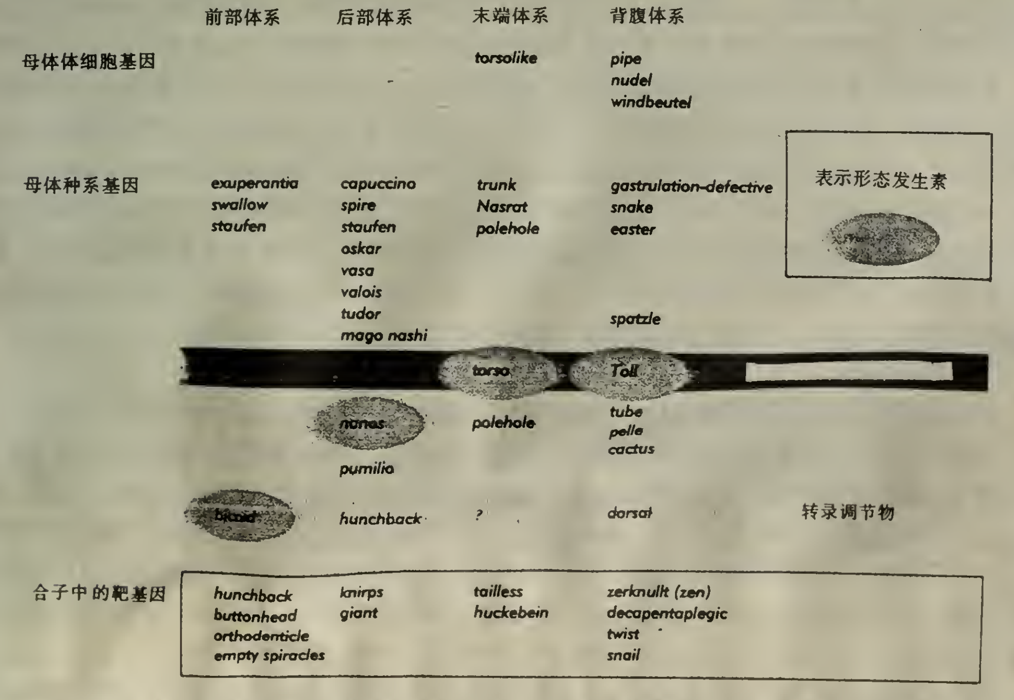


图 13-54　4 个母体体系，每个都是由其产物作用于卵外侧的基因开始的，然后途径转到卵内，在卵内每个途径的局部产物都是形态原，它们或者是个受体，或者是个基因表达的调节物

途径最后的成分是个转录因子，它作用于合子的靶物质，这些靶物质负责下一阶段的发育。

(morphogen)。形态发生素的定义是，它是个蛋白质，它的局部浓度（或活性）使其周围区域采取一种特定的结构或结局。形态发生素有的是转录的调节因子，有的是使一个转录因子活化。4 个体系中的 3 个与前-后轴有关，一个与背-腹轴有关。简述如下：

（1）前部体系（anterior system） 负责头和胸部的发育。此体系的形态发生素是 Bicoid 蛋白，它是个转录调节因子，调控基因 *hunchback*（大概还有其它分节的和同源异型基因）的表达。*bicoid* RNA 是在滋卵细胞中生成的，并依靠母体的生殖系产物转运至卵母细胞中，并局限在卵的前端。

（2）后部体系（posterior system） 负责腹部的分节。其最初过程的本质尚不清楚，是由许多产物的作用使 *nanos* 产物在卵母细胞中局部化。*nanos* 产物是形态发生素，它局部阻遏 *hunchback* 的表达（通过 mRNA 翻译的调控）。

（3）末端体系（terminal system） 负责卵的不分节末端的特异结构（头部的原头区（acron）和尾部的尾节（telson））的发育。已知此体系的最初事件发生在滤泡细胞，这些事件导致 *torso* 编码的跨膜受体局部活化。此途径的最终产物尚待鉴定。

（4）第 4 个体系是负责背-腹发育的 此途径是由来自滤泡细胞的一个信号在卵的腹部启动的。此信号由 *Toll* 编码的跨膜受体转导，引起 *dorsal* 产生的转录因子的活性梯度（用控制它在细胞中的定位做到的）。

已鉴定出约 30 个参与格局形成的母体基因，而且看来也不会比这个数目再多很多了。4 条途径的所有成分都是母体的，可见，建立最初格局的所有体系都取决于受精前发生的事件。两个体轴（body axis）是各自独立建立的。影响极性的突变可使后区发育成像前部的结构，或者使腹区发育成背的样子，但影响一个轴的突变不影响另外的轴。在前-后轴上，前部体系和后部体系提供对立的梯度，控制着体节的发育。末端体系和背-腹体系的运转不依赖于其它体系。

发育所需的物质浓缩于局部的事实是用营救法（rescue）表明的。此法是将由野生型胚胎中取出的物质注射给早期发育有缺陷的突变胚胎。如果突变的胚胎因此而发育正常，则可知突变使它缺乏了在野生型胚胎中存在的某物质。例如用此技术曾鉴定了前部的决定因素。已知 *bicoid* 基因的突变体不发育头部，但此缺陷可通过给突变的卵注射由野生型胚胎前端取得的胞浆而纠正。而且，注射野生型前部胞浆还可使突变胚胎的其它部位发育出前部结构来。营救的程度还取决于注射的野生型胞浆的量。注射物中的活性成分可被提纯和鉴定，在 *bicoid* 的例子中此活性成分是 *bicoid* mRNA，注射提纯的 *bicoid* mRNA 可代替前部胞浆。

（一）前部体系

bicoid 基因是在滋卵细胞中转录的，产生的 mRNA 通过胞质桥转运到卵母细胞中。此 mRNA 定位于胚胎的前部末端，但在卵子发生期（oogenesis）它不被翻译。翻译是在受精后立即开始的，然后此蛋白质沿胚胎建立起一个梯度，如图 13-55 所示，此梯度可能是由于 Bicoid 蛋白由前部末端的源头扩散产生的，此梯度建立于 7 次分裂之前，并保持稳定至囊胚层阶段之后。已发现前部发育所需的其它基因，*swa* 和 *exu* 的产物也具有相同方向的梯度。这些基因产物的功能可能与 *bicoid* mRNA 转运入卵母细胞有关，或者与使 *bicoid* mRNA 不由前部末端向外扩散有关。*bicoid* mRNA 的定位于前部末端依赖于其 3′端非翻译区的序列。

Bicoid 梯度建立起来的后果是什么？用改变母体功能基因的拷贝数而使此梯度增高或降低的实验表明，减弱此梯度引起前部的体节（segment）发育得更像后部的体节，而增强此梯度则使前部样（anterior-like）的结构沿胚胎延伸得更远。可见 Bicoid 蛋白是个形态发生素，依靠

它的浓度决定着胚胎中前部-后部的位置。

胚胎前部细胞的结局是由它们所处位置的Bicoid蛋白的浓度决定的。Bicoid蛋白是个转录因子，它特异地启动一些基因的转录，这些基因启动后再进一步调节另外一些基因的开关，从而决定着前部结构的形成。由此可见，在前部发育中Bicoid起着指导性作用。已知Bicoid直接启动的一个主要基因是*hunchback*，其它基因尚待鉴定。Bicoid启动*hunchback*的转录是剂量依赖型的，即在Bicoid蛋白的浓度高于一个特定阈值时*hunchback*才被活化，这就使得Bicoid蛋白的浓度梯度能够影响处于不同位置的基因是否表达。在胚胎发育中，就是以这种方式使形态发生素(如Bicoid蛋白)的量的差异转变成了(细胞结构)在质上不同的状态。关于Bicoid启动*hunchback*的后果将在后面讨论。

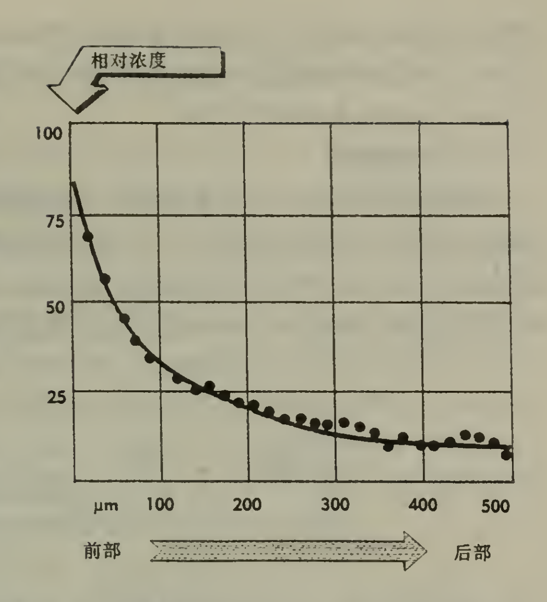


图13-55　在果蝇发育过程中，Bicoid蛋白形成一个梯度。此梯度沿500 μm长的卵延伸约200 μm
由于蛋白质含量的测定方法还不够精确，梯度的实际情况可能与此图有所不同。

（二）后部体系

后部的发育依赖于一大组基因的表达。这些基因中任何一个发生突变的雌蝇产生的胚胎，其头和胸节都正常发育，但缺乏腹部。后部体系的起始事件与前部体系的基本性质相似，即一个母体的mRNA，*nanos* mRNA定位于后部的极点处(posterior pole)，其翻译产物Nanos蛋白是个形态发生素，它由其合成位点扩散，沿腹区形成一个梯度。确定Nanos蛋白为形态发生素是因为注射提纯的*nanos* mRNA能营救任何其它后部基因发生突变的胚胎，而且将*nanos* mRNA注射给胚胎的其它部位，能引起该处形成腹部结构。其它后部基因的作用则可能与*nanos* mRNA的转运和定位有关。*nanos* mRNA的定位依赖于其3′非翻译区的序列。但它的定位比*bicoid* mRNA的定位要复杂得多，因为*nanos* mRNA是由滋卵细胞产生的，它由卵母细胞的前端进入，需要转运通过卵母细胞的全长，才能达到并定位于后部极点。

*bicoid*和*nanos*都作用于*hunchback*基因的表达。*hunchback*编码一个转录的阻遏物，前部结构(胸区)的形成需要它的存在，后部结构的发育则不需要它。在卵母生成期它被转录产生mRNA，此mRNA均匀地分布在卵中。受精后，在前部区域中Bicoid的梯度活化*hunchback* mRNA的合成；在后部区域*nanos*阻止*hunchback* mRNA的翻译，其结果是使*hunchback* mRNA降解。因此，前部体系和后部体系共同使得卵的前半部*hunchback*水平升高，而在后半部则将它除掉。后面即将讨论*hunchback*这种分布的意义在于它调控着一些基因。已知它直接阻遏基因*knirps*和大概还有*giant*，这些基因是形成腹部结构所必需的。可见*hunchback*的主要作用是通过抑制*knirps*和*giant*在前部区域的表达，从而阻止了在此区域形成腹部结构。

后部体系除了如上所述负责腹部的发育外，还负责极细胞(polar cells)的形成，极细胞将生成生殖细胞。这两种事件的形态发生素都存在于极质中。其中之一是*nanos*，如前所述它决定腹部的发育；另一个形态发生素尚未鉴定出来，它调控极细胞的形成。由于除了*nanos*和*pumilio*之外，其余所有后部基因都是两个过程所必需的，可见后部体系在开始时是一个途径，

到了一定阶段发生分支。分支后由 *nanos* 和 *pumilio* 确定腹部发育。Nanos 蛋白是形态发生素，*pumilio* 的功能可能是其产物协助 Nanos 蛋白的扩散；活化或稳定 Nanos 蛋白。而极细胞形成分支的形态发生素则尚不清楚。

（三）背腹体系

果蝇胚胎背-腹样式的发育需要 11 个母体基因，它们的功能是在由受精到细胞型囊胚层之前的期间建立起背-腹轴（见图 13-54）。此体系对腹部结构[其中包括中胚层(mesoderm)和神经外胚层(neuroectoderm)]的发育是必需的。如图 13-56 所示，这些基因中的任一基因的突变体都在其腹侧有背部结构。但将野生型的胞质注射到突变的胚胎中可挽救其缺陷，使腹部结构能够发育。

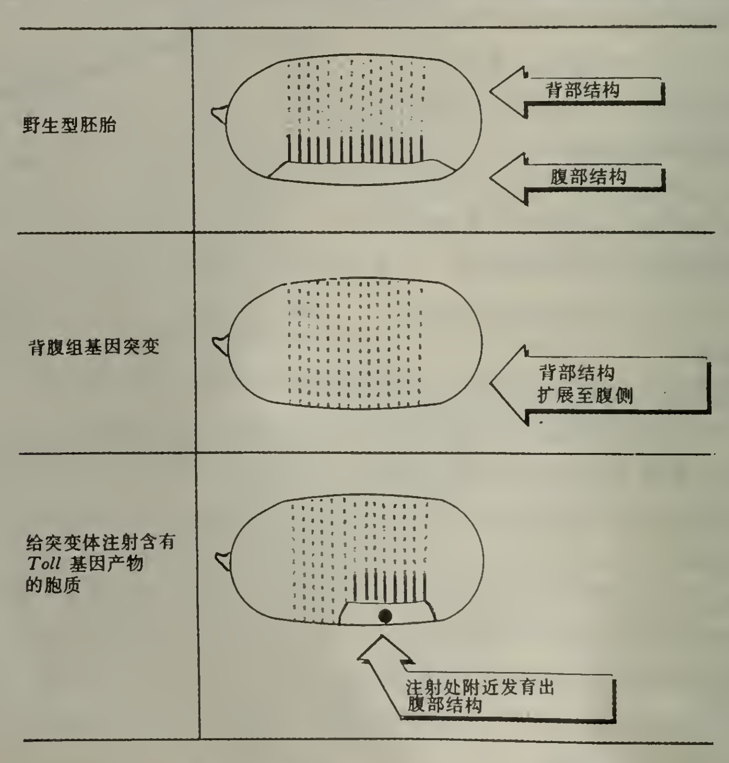


图 13-56　野生型果蝇胚胎有明显不同的背部和腹部结构

背-腹组基因的突变阻止了腹部结构的出现，胚胎的腹侧背部化了。注射含有 *Toll* 基因产物的胞质可恢复腹部结构。

在此途径中卵母细胞和滤泡细胞之间发生着复杂的相互作用。卵母细胞中某些基因的功能是滤泡细胞正常发育所必需的，由于这些基因的突变而使滤泡细胞有缺陷时，则背-腹的发育不正常。此现象的发生是因为在滤泡细胞中有 3 个基因（*pipe*，*nudel* 和 *windbeutel*），它们的功能是传送一个信号给卵母细胞（卵在受精前不接受此信号。）此信号的本质虽然尚不清楚，但已知它引起在卵的表面上发生一系列的蛋白质裂解作用：*snake* 的产物使 *easter* 的产物裂解而活化，活化的 *easter* 产物又使 *spatzle* 的产物裂解而活化，活化的 *spatzle* 产物则是 *Toll* 基因编码的受体的配体(ligand)，*Toll* 是此途径在卵母细胞中起作用的第一个成分。营救实验表明，*Toll* 是将信号转入卵母细胞内的关键，途径中的其它基因则起辅助作用。

Toll 蛋白的作用是诱导背-腹结构的形成，它仅在腹侧有活性而在背部没有，从而形成背-腹梯度。然而实验证明，在胚胎的所有部位都有 Toll 蛋白，那么为什么它仅在适当的局部诱导

出腹部结构？这是因为它的功能需要结合上配体，而此配体(Spatzle 蛋白裂解产物)仅在卵周隙(perivitelline space)的腹侧产生，此配体或是由于不能扩散至远处，或是由于与 Toll 结合得很快，其结果是使 Toll 仅在胚胎的腹侧被活化。

Toll 蛋白活性的本质目前尚不清楚，但已知它是以某种方式经由 *tube* 和 *pelle* 而起作用的。*tube* 的功能尚不清楚。*pelle* 编码一个激酶，此激酶的靶子大概是 *cactus* 的产物，此产物调控着 *dorsal* 编码的转录因子。在胞质中 Cactus 蛋白和 Dorsal 蛋白形成复合体，使 *dorsal* 不能进入核。当 Cactus 蛋白被 Pelle 蛋白磷酸化后，它将 Dorsal 蛋白释出，Dorsal 蛋白即可进入核中。由此可见，Toll 活化的结果是在胚胎中由腹侧至背侧建立起 *dorsal* 在核中活性的梯度。在胚胎的腹部 Dorsal 蛋白进入核，在背侧它仍停留在胞质中。此梯度是在合胞体囊胚层(syncytial blastoderm)阶段建立的，实验证明，它在过渡到细胞型囊胚层期间梯度变得更陡。在此过程中，胚胎中 Dorsal 蛋白的总量没有改变，只是由于它在核和胞质中的再分配而建立起了梯度。

Dorsal 蛋白活化基因 *twist* 和 *snail* 是腹部结构发育所必需的。Dorsal 蛋白又阻遏背部发育所必需的基因 *dpp* 和 *zen*，其结果是这些基因仅在背部表达，发育出背部结构。

(四)末端体系

末端体系启动的方式与背-腹体系相似。受精后，通过翻译母体的 *torso* mRNA 产生一种跨膜受体。此受体定位于整个胚胎中，但在极部它被局部产生的一种细胞外配体所活化。Torso 蛋白具有激酶活性，它启动一个级联使 *tailless* 和 *huckebein* mRNA 局部表达，这两种 mRNA编码调控转录的因子。

(五)小结

现将每种体系的调节物在发育的每个阶段的分布情况总结于图 13-57。最初的不对称是

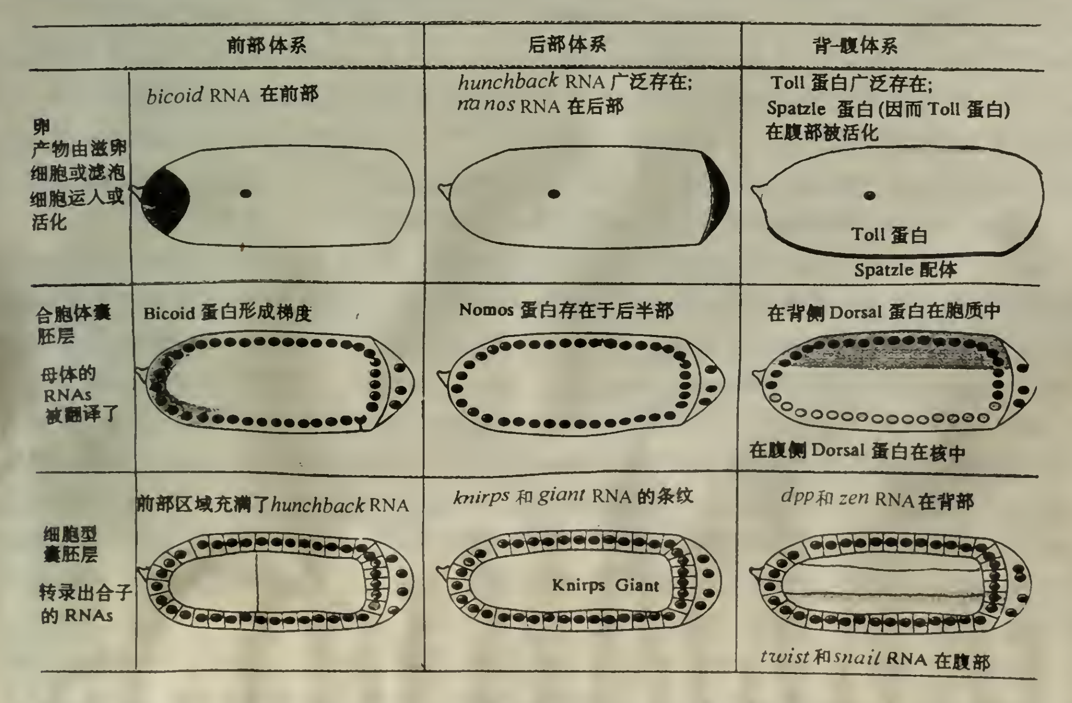


图 13-57 在每个轴的决定体系中，卵中的局部产物使得在合胞体囊胚层阶段其它母体 RNAs 或蛋白质较大范围的定位，在细胞型囊胚层期合子的 RNA 在不同带中生成

用两种机制建立的。对于前-后轴来说是将一种 mRNA(前部体系是 *bicoid*,后部体系是 *nanos*)定位在卵的一端,定位依赖于 mRNA 的 3′端序列与母体蛋白质的相互作用。在背-腹体系和末端体系中,则是一个受体蛋白在特异的局部被活化,这是因为仅在此局部有其配体的结果。所有这些相互作用都依赖于母体基因表达产生的 RNAs 和/或蛋白质。

上述物质的定位于局部引起一个形态发生素梯度的产生。有的是量的梯度(Bicoid 蛋白),有的是由于在核和胞浆中的不同分配(Dorsal 蛋白),有的则是定位于一个较宽的局部区域内(Nanos 蛋白)。每个体系的形态发生素活动范围约为卵的 50%。形态发生素是母体 mRNAs 翻译产生的,所以直到此阶段发育是依赖母体基因的。

建立前-后和背-腹梯度是确定胚胎的方位和空间结构的第一步。在母体基因的指导下,跨越共同的胞质形成梯度,它影响定位于其中的核的行为。下一步是分隔的区域的发育,不同分隔区将生出身体的不同部位,这需要合子的基因组的表达,此时变为活动的基因座(loci)称为合子基因(zygotic genes)。参与这一阶段的基因是用分节(segmentation)突变鉴定出来的。

三、囊胚层阶段形成的分区决定着细胞的结局

由于如上所述,在母体基因的作用下,沿前-后轴建立起来了梯度,从而使不同基因在不同的区域表达,其结果使胚胎形成许多节,每个节分化为成虫的特定形态。这些节是怎样形成的?为了说明这个问题,我们首先考察一下由胚胎到成虫各相应部位之间的关系,如图 13-58 所示。从图中可以看出,成虫是由一系列的可以清楚区分的体节(segment)所组成,而幼虫则有一系列用沟(groove)分开的相应的节。我们将主要讨论 3 个胸部(T)和 8 个腹部(A)的节,因为对它们的发育知道得最多。节的格局是在囊胚层期决定的。在囊胚层期胚胎的主要物质被分开为一系列交替的前部(A)和后部(P)分隔区。一个体节是由一个 A 分隔区和其后的一个 P 分隔区所组成,例如,节 A3 是由分隔区 A3A 和 A3P 组成的。而在节出现之前,即在原肠胚(gastrula)期可看到副节(parasegment),在 5～6 h 的胚胎表面有浅沟将相邻的副节分开。如图中所示,每个副节都是由一个 P 分隔区和其后的一个 A 分隔区所组成,例如,副节 4 是由分隔区 T1P 和 T2A 所组成。当发育至 9 h 左右形成节时,沟加深并移动,因此每个节的边边代表着一个副节的中心,即节的前半部来自一个副节、后半部则来自下一个副节。

由上述可见,在果蝇的发育过程中是在胚胎发生期形成分隔区,然后再由这些分隔区形成副节和节,那么这些分隔区是怎样建立的?突变的实验表明,大约有 30 个基因参与节的生成。根据这些基因突变产生的后果可把它们分为 3 类,即间隔基因(gap gene);配对-界定基因(pair-rule gene)和节极性基因(segment polarity gene),如表 13-5 中所列。这些基因统称为分节基因(segmentation genes)。在发育过程中,这 3 组基因逐次地表达,如图 13-58 所示,它们规定着卵的越来越狭窄的区域。从图 13-59 中可以看出,首先是母体的基因由前部末端和后部末端建立起梯度。这种梯度活化或阻遏间隔基因。间隔基因是受精后最早(第十一次核分裂后)转录的基因,它们将胚胎分为 4 个宽的区域。间隔基因调节配对-界定基因。配对-界定基因转录得稍晚一些,它们被限定在一对节的区域内。配对-界定基因再调节节极性基因,节极性基因在第 13 次核分裂期间表达,它们的区域大小是每个节。

许多母体基因、间隔基因和配对-界定基因是转录的调节因子,它们的产物活化或阻遏转录。在有些情况下,一种蛋白质可活化一些基因、阻遏另一些基因,这取决于它的含量高低或与其它物质的关系。任一组的基因既彼此调节,又调节下一组的基因。只有节极性基因的作用方

式不同,例如,一种蛋白质可由一个细胞中分泌出来去影响它邻近的细胞。由此可见,在每一阶段都是用少数母体的、间隔的和配对-界定的蛋白质联合起来发挥作用的方式,以规定在胚胎的特定区域基因表达的样式 。

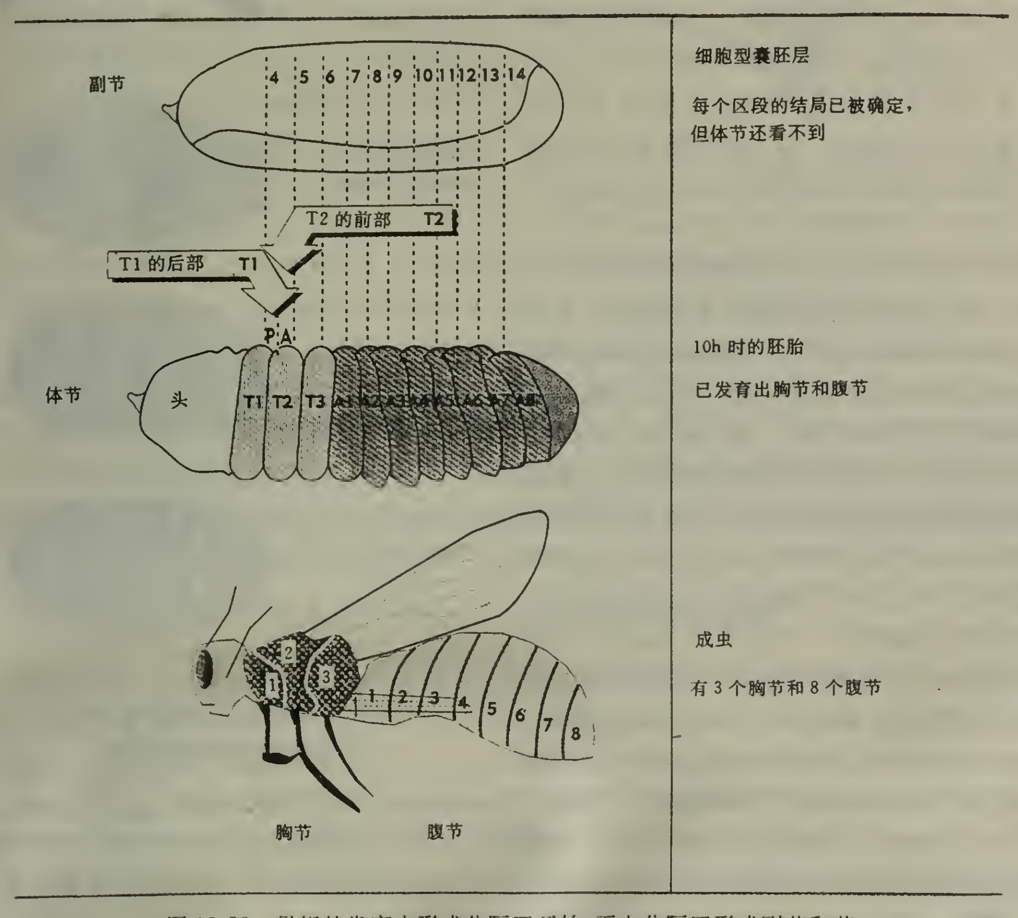


图 13-58 果蝇的发育由形成分隔区开始,再由分隔区形成副节和节

表 13-5 分节基因的分类及其表达时间

基因	最早表达的时间	基因	最早表达的时间
间隔基因		节极性基因	
hunchback		*engrailed*	
kruppel		*wingless*	
knirps	<11	*gooseberry*	
giant	核分裂	*cubitus interrupts*	
tailless		*patched*	
配对-界定基因		*hedgehog*	13
runt		*dishevelled*	核分裂
hairy		*costal2*	
ftz	11～12	*fused*	
even skipped	核分裂		
paired			
odd paired			
sloppy paired			
odd skipped			

这种调控的机制非常复杂，其中的很多细节目前还不清楚。为了说明这种调控的基本原理，现用图 13-59 中间隔基因规定的 4 个区的产生方式加以说明。已知最前端的带是由 Hunchback 蛋白组成的，它由 *hunchback* mRNA 翻译生成，而此 mRNA 的合成是由 *bicoid* 活化的。第二条带由 Kruppel 蛋白组成，*kruppel* 基因的转录由 Hunchback 蛋白活化。后两条带分别由 Knirps 和 Giant 蛋白所组成，它们的基因的转录受到 Hunchback 蛋白的阻遏，但在这些区内 *nanos* 阻止了 Hunchback 的表达，因而在胚胎的后部 *knirps* 和 *giant* 表达了。图 13-60 表示出由间隔基因的产物间隔蛋白(gap protein)相互作用而形成 4 个宽带的情况(由图中可见实际情况比我们说的还要复杂些)，这是用突变实验得出的结果。在其中 *hunchback* 起着特殊重要的作用，它在胚胎前部一个宽的区域中表达，到中部降低了。高水平的 *hunchback* 阻遏 *kruppel*，以此规定 *kruppel* 表达的前方界限。在副节 3 处随着 *hunchback* 的降低 *kruppel* 的表达升高。然而 *kruppel* 的表达又需要一定水平的 *hunchback*，所以在副节 5 附近，当 *hunchback* 的水平进一步降低时，*kruppel* 关闭了。*hunchback* 以相同的方式调控着 *giant* 的表达。*knirps* 则在没有 *hunchback* 处表达。已证明间隔蛋白间相互作用以调控基因表达的实际情况比上述要复杂得多。实际上沿着卵的前部到后部的每个特定位置，都被各种间隔蛋白的水平确定了。

母体基因由前端和后端建立起梯度

间隔基因在卵中分隔在 4 条宽带中

配对-界定基因建立 7 个条纹

节极性基因建立 14 个条纹

图 13-59　母体的和分节的基因逐次作用在胚胎的越来越小的区域

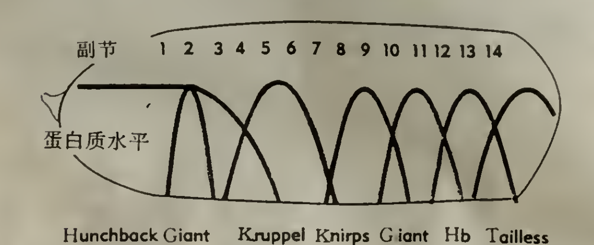


图 13-60　间隔基因的表达将胚胎划分成不同的区域曲线示出各种蛋白质的水平。

所有间隔蛋白都是转录的调节物，它们除了彼此相互调节外，还调控着在下一阶段起作用的配对-界定基因的表达。由于在相当于不同体节处各种间隔蛋白的浓度有不同的搭配，而对于某一配对-界定基因来说，这些间隔蛋白有的是激活物，有的是阻遏物，从而使配对-界定基因仅在某些副节中表达。例如，使 *ftz* 基因仅在偶数副节中表达，而 *eve* 基因仅在奇数副节中表达。以 *eve* 基因在副节 3 中的表达为例可具体说明这种调控的原理。在副节 3 中用分布范围较广的 Hunchback 蛋白和 Bicoid 蛋白活化 *eve* 基因使之表达，但其表达边界用两个阻遏物限定：其前边边界用 Giant 蛋白限定，后边边界用 Kruppel 蛋白限定，如图 13-61 所示。已知造成这种结果的原因是由于在 *eve* 的启动子中有很多应答元件，其中有些结合上 Hunchback 或

Bicoid 蛋白而被激活，有些则结合上 Giant 或 Kruppel 蛋白而被阻遏。

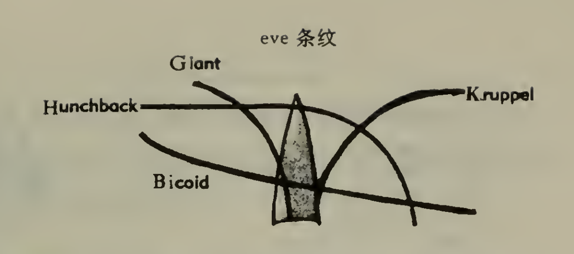


图 13-61　调控 eve 在副节 3 中表达的原理示意图
图中曲线示出在 eve 条纹附近几种蛋白质的水平。

配对-界定基因调节节极性基因的表达，但其调节机制目前还不了解。已知节极性基因在每个体节中表达，而且表达的区域是非常严格的。这样就使每个节中的细胞分化成为不同的细胞，并发育成不同的结局。而且节极性基因还将每个节分隔成为前部(A)分隔区和后部(P)分隔区，已知这是由于 *engrailed* 基因表达的结果。*engrailed* 基因仅在 P 分隔区中表达，不在 A 分隔区中表达，从而把每个节分成为 A 分隔区和 P 分隔区。当此基因发生突变时，节的 A 分隔区和 P 分隔区就区分不开了。

在这里有两点需要加以说明：①在前面的叙述中强调了母体基因调控间隔基因的表达，间隔基因调控配对-界定基因，配对-界定基因再调控节极性基因的基因级联调控方式，在总体上说这无疑是对的，但实际上并不很严格，它们有着复杂的相互作用。例如，*bicoid* 是母体基因，但其产物直接调控着配对-界定基因 *eve* 的表达；②许多分节基因的产物是转录因子，但并非所有的都是。已知节极性基因的产物中有分泌蛋白质、跨膜蛋白质、激酶和细胞骨架蛋白等，它们如何发挥作用的机制目前还不清楚，但可以肯定它们是以不同方式发挥作用的。

四、同源异型基因调控胚胎的基本形态结构

如上所述，分节基因调控着每个体节内的样式，包括其极性。这样每个体节中的细胞是不同的，它们将发育出不同的形态结构。然而某些突变可把身体一个部位转化(transform)为另一种部位，这种突变叫做同源异型突变(homeotic mutations)。例如，触角-足(Antenna-pedia，*Antp*)突变使一个触角变成了一条腿。这说明正常的 $Antp^{+}$基因确定在正确的位置形成腿。我们把用这种突变方法鉴定出来的基因称为同源异型基因(homeotic genes)。一般说来，同源异型基因负担着决定每个节的独特分化程序，它们调控着胚胎的形态构筑方案。同源异型基因是在胚胎发生期活动的，它们的表达依赖于此前分节基因的表达，因此可把同源异型基因的作用看成是对分节基因建立起来的信号格局的整合。尽管同源异型基因的突变可引起诸如在触角位置发育出腿，或者在眼睛处发育出翅，然而它们并不改变分节基因等基因已经建立起来的信号格局。同源异型基因是通过开动特定位置的一套基因来修饰该位置细胞的结局，而细胞的结局是已经由节极性基因等确定了的。

在果蝇中同源异型基因成簇地存在于两个主要复合座(complex loci)中。触角足复合座(antennapedia complex，ANT-C)调节头和前部的胸节；双胸复合座(bithorax complex，BX-C)调节后部的胸节和腹节。

图 13-62 示出了 ANT-C 的结构。ANT-C 中含有数个同源异型基因，其中包括 *labial* (*lab*)，*proboscipedia* (*pb*)，*Deformed*(*Dfd*)，*Sex combs reduced*(*Scr*)和 *Antennapedia* (*Antp*)。

这些同源异型基因成簇地位于一个约 350 kb 的区域中，但其中散布有几个其它的基因。果蝇最前面的几个部位(副节 1～4)的特性是由 ANT-C 决定的，图中指明了 ANT-C 的同源异型基因分别在胚胎中表达的部位。这些基因在 ANT-C 中由左向右的次序也就是它们在胚胎中由前向后表达的位次。例如，*labial* 在最前面，它影响头部，最后面的是 *Antp*，它影响节 T2～T3。

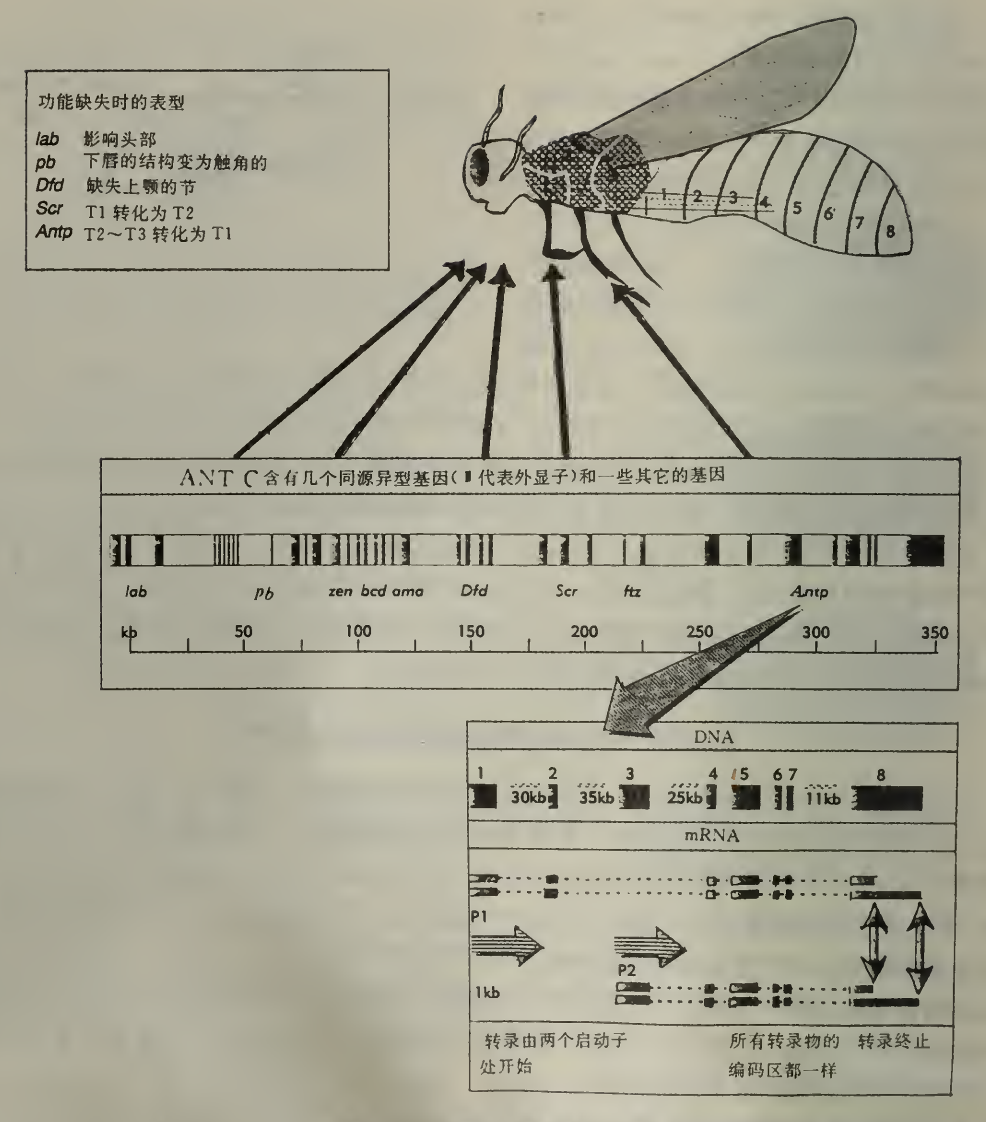


图 13-62　ANT-C 复合座的同源异型基因赋予果蝇前部各节的特性

在同源异型基因中间还散布有其它基因。*Antp* 基因很大，它有不同的表达方式。

图 13-62 中还示出了 *Antp* 基因的结构。此基因约 103 kb 长，有 8 个外显子和很大的内含子。实际编码蛋白质的部分仅为基因 DNA 的 1%。它有两个启动子，一个位于外显子 1 的上游，另一个位于外显子 3 的上游，二者相距约 70 kb。

转录的终止位点也有两个。然而无论转录从何处开始和何处终止，编码的蛋白质是相同的。

BX-C 复合座是用几种同源异型突变鉴定出来的，这些突变影响了胸部发育，引起了腹部的形态发育。当整个复合座缺失时，昆虫在胚胎发育的后期死亡。然而复合座内发生突变时可以存活，但一些节的表型发生改变。图 13-63 是个极罕见的例子。图中果蝇由于 3 突变（triple mutation），使 T3A［它带有平衡棒（haltere），截短了的翅］转变为 T2 的组织型（它带有翅），从

而使果蝇有 4 个翅，而不是通常的两个翅。

BX-C 的基因调控的范围由 T2 到 A8，可见，BX-C 复合座与果蝇身体的大部分发育有关。与 ANT-C 一样，BX-C 的突变与其影响的节的关系也是基因图谱的次序与节的次序相同，即沿基因图谱由左向右依次突变，则由前向后依次影响身体各体节。

同源异型基因是怎样影响每个体节的形态的？如前所述，大多数同源异型基因是在有限的空间表达的，每个基因的表达空间相当于一个副节。许多同源异型基因编码的是转录因子，它们的表达产物各自调控着其靶基因的表达，而这些靶基因的表达则表现出每个节的表型。然而遗憾的是对于这种靶基因目前我们还一无所知。



图 13-63　在 BX-C 复合座的 *abx*，*bx* 和 *pbx*3 处突变，产生一支四翅果蝇

同源异型基因本身的表达是受到严格调控的，已知这种调控是很复杂的。它们的表达既受到在它们发挥作用之前的分节基因表达情况的调控，又受到彼此之间的调控。已知这些基因具有很多调控元件，这些元件分布在一个很长的区域内，使得它们具有了很复杂的调控方式。因此，一个同源异型基因的突变引起的最后表现不仅取决于该基因功能的丢失，而且还取决于这种丢失所引起的其它同源异型基因如何改变的后果。因此，增加了对同源异型基因突变的后果进行分析的复杂性，而这种分析一直是研究有关问题的重要手段。

现在我们可以概括一下果蝇发育的过程。早期胚胎是个合胞体，在其中许多核处于同一胞质中。此时由母体基因的产物建立起前-后轴和背-腹轴梯度，从而能够按照每个核所处位置的不同控制每个核的功能。到了细胞型囊胚层阶段，合子的 RNAs 开始合成，此时胚胎的发育变为依赖它自己的基因。细胞是在囊胚层阶段形成的，形成后发生一个转录调节物的级联，由分节基因产物的相互作用决定每个体节中基因的独特表达。同源异型基因在每个体节中的独特表达决定了每个体节的最后表型。

五、同源异型框是同源异型基因中的一个共同的编码基序

其它生物是怎样发育的？例如，在发育的早期是怎样由一个单一的细胞经过次数不多的分裂后，就使产生的子细胞分化为性质不同的细胞的？其它生物产生这种分化的方式可能与果蝇的不同。例如，哺乳动物的卵可能是均一的，而不是像前面所说果蝇的卵在开始时其胞质成分的分布就是不对称的。然而同源异型框（homeobox）的发现及其研究得知，控制发育的分子机制在进化中有着高度的保守性。

同源异型框是 David Hogness 和 Walter Gehring 在果蝇的 ANT-C 和 BX-C 中首先发现的。他们用杂交试验揭示出在这些簇中的许多基因中含有一个共同的序列，并证明此一再出现的是个 180 bp 的序列，称之为同源异型框。它编码一个 60 氨基酸残基的多肽，称为同源异型域（homeodomain），见图 13-64。同源异型域的碱性氨基酸残基含量很高，提示它是与 DNA 结合的。事实是含有同源异型域的蛋白质定位于细胞核中，并结合在特异的 DNA 序列上。

诚然，在由酵母到人类的所有真核生物中，同源异型域都是基因表达的调节物。用核磁共

振和 X 射线结晶学研究果蝇和酵母的同源异型域表明，它们有一个伸展的氨基-末端臂和 3 个 α 螺旋(图 13-64 中示出了其部位)。螺旋 1 和 2 是彼此反平行的，它们与螺旋 3 近于垂直。其氨基-末端臂适合进入 DNA 的小沟，而螺旋 3 则适合进入大沟。同源异型域以单体结合在 DNA 的特异位点上，其螺旋 2 和螺旋 3 安排为螺旋-转角-螺旋样式与 DNA 结合。

螺旋 1　　螺旋 2　　螺旋 3

A SKRGRTAYTRPQLVELEKEFHFNRYLMRPRRVEMANLLCLTERQIKIWFQNRRMKYKKDN
B RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHVLCLTERQIKIWFQNRRMKWKKEN
C RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN
D KPYRGHRFTKENVRILESWFAKNPYLDTKGLENLMKNTSLSRIQIKNWVSNRRRKEKTIT

图 13-64　同源异型域是由调控发育的基因编码的一再出现的基序

哺乳动物、两栖类、昆虫和酵母中的序列是相似的。图中比较的同源序列为：(A)小鼠 *MO*-10 基因；(B)蛙 *MM*3 基因；(C)果蝇 *Antp* 基因和(D)酵母 *MATαZ* 基因。

在进化中同源异型域的保守性是突出的。脊椎动物的同源异型框基因(homeo box gene)叫做 *Hox* 基因，它们与其在果蝇中的同系物(homologs)是相似的。在小鼠和人中同源异型框基因被安排在 4 个簇中，即 *Hox*1，*Hox*2，*Hox*3 和 *Hox*4，它们存在于不同染色体中。由小鼠 *Hox*1.1 基因编码的同源异型域与果蝇 *Antp* 编码的仅差 1 个氨基酸残基。而且哺乳动物的 *Hox* 基因在果蝇中也起作用。更值得注意的是，在果蝇和哺乳动物二者中同源异型框基因的顺序都是精确的相当于它们沿胚胎的前-后轴发生作用的次序(图 13-65)。规定头部发育的同源异型框基

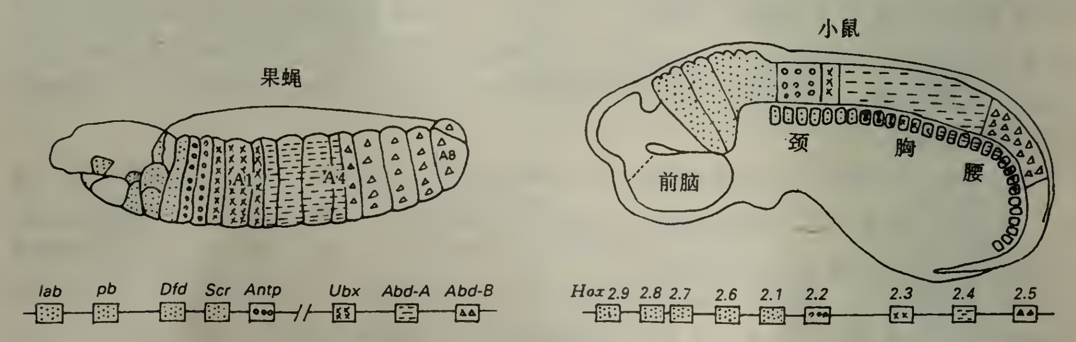


图 13-65　在进化中，同源异型框基因的安排和它们编码的蛋白质的序列是高度保守的

左图是果蝇的 ANT-C 和 BX-C 中基因的次序以及每个基因在果蝇胚胎中表达的位置；右图是小鼠 *Hox*2 基因相应的基因以及它们表达的位置

用不同点线表示出果蝇和小鼠间相应的基因以及它们表达的情况。

引自 E. B. Lewis. J. Amer. Med. Assoc. 1992. 267:1 524.

因在簇的一端，而指导后部结构(在果蝇为腹部，在哺乳动物为腰椎)形成的在另一端。很明显，划定体节的特异化这个发育的基本机制在进化中已经保留了 6 亿多年。现在我们已在开始破译包含在 DNA 序列中的一维信息是怎样转变为多细胞生物的复杂而美丽的三维形式的。

参 考 文 献

1. Lehninger, A. L., Nelson, D. L. and Cox, M. M. 1993.
2. Principles of Biochemistry (2nd ed.). Worth Publishers, (Chapters 25 and 27.).
3. Stryer, L. Biochemistry (4th ed.). W. H. Freeman and Company, (Chapters 37.) 1995.
4. Lewin, B. Genes V. Oxford University Press, (Chapters 13, 29, 30, 31, 32 and 38.) 1994.

(齐顺章)

第十四章　分子生物学技术

第一节　引　　言

分子生物学的发展是基于我们对生物大分子的了解，尤其是DNA和RNA。通过对前面章节的学习，我们知道，DNA和RNA是生命体中遗传信息的载体，生物学中的许多秘密就隐藏于DNA的核苷酸序列中。但是要得到一段较长的DNA序列，在70年代初期以前似乎还只是一个幻想，更不用说现在随处可见的任意改变DNA的序列这样一类实验。然而，随后的技术上的进步迅速改变了这一状况。首先，能够用酶在特定位点切割DNA，并将其断裂成可重复的不同大小的片段，这些现在称之为限制性内切酶的应用，使两种非常重要的技术，即DNA克隆和DNA序列分析得到了极大发展。

两个DNA分子可以在酶的催化下连接起来。因此，任何DNA的限制性片段都能插入到一个载体DNA分子中（通常是质粒DNA分子），这样，就形成了重组DNA分子，将重组DNA分子引入到合适的细胞群体中之后（常见为细菌），带有特定重组DNA分子的细胞可被挑选出来。这一过程称之为目的DNA片段的克隆。由此，我们可以制备出无限量的特定DNA分子。随着技术的发展，现在能够利用化学方法在机器上自动合成长度达100碱基的DNA寡聚核苷酸。因此，人们不仅能够制造出含有天然DNA的重组DNA分子，而且也能够将经过突变的或人工合成的DNA分子插入到载体分子中。

DNA序列的快速分析法是70年代末发展起来的。利用限制性内切酶，人们可以完全重复地把从某一生物体中得到的DNA切割成大小不同的片段。这些片段在原来的DNA分子中的排列顺序可以通过实验得以确定。这时，确定一个长度为500 bp的DNA片段的顺序也已成为现实。因此，要分析一个长达10 kb的DNA分子的序列也已没有任何困难。也就是说，任何DNA分子都可进行分离和序列分析。这为人们了解复杂的真核生物基因组结构开辟了道路。依靠计算机的帮助，我们现在可以自动分析DNA的序列，并将序列资料进行贮存、比较和分析。可以预见，人们还将在不远的将来了解整个人类基因组的顺序。

现在的分子生物学技术，还能够把任何DNA，无论是天然的还是经过改造的，甚至是完全人工合成的，送回到细胞中，以鉴定其生物学活性。如果一段编码蛋白质的基因被送回到细菌或其它细胞中时，它将指导所编码的天然的或突变的蛋白质的合成。

分子生物学的发展，已经产生了一整套分子生物学技术。几乎在一夜之间，这一技术已经成了研究许多基本生物学问题的重要手段。通过对DNA或RNA的提取，基因的分离，核苷酸序列的测定，人们将了解到基因及其所编码的蛋白质的结构，进而对其功能也有所认识。通过对基因的表达的研究，我们能了解基因的表达调节过程，进一步认识生命活动的规律。通过对基因进行突变和改造，我们将深入了解基因及其产物的结构和功能的关系，基因在生长发育过程中的作用等。总之分子生物学技术在生命科学各个方面的渗透，使我们掌握有效的手段，为

进一步认识生命活动的本质，更好地利用其规律为人类服务提供了帮助。本章将简要介绍分子生物学技术的基本原理。

第二节 放射性同位素技术

放射性同位素作为一种示踪物已经被广泛用于分子生物学技术中。放射性标记示踪实验简化了许多生物化学的测定方法，具有用量少、灵敏度高的优点，而且还具备一定的特异性。常用的放射性同位素为^{3}H，^{14}C，^{32}P，^{33}P，^{35}S，^{125}I 等，其中以^{32}P 和^{35}S 应用最多。

在被标记的分子中，至少应有一个原子是放射性同位素原子，这一原子并不改变被标记分子的化学性质。也就是说，如果被标记的分子是一个酶，当它执行代谢功能时，如：催化蛋白质，DNA，RNA 的合成与未标记的相同酶分子无任何区别，但由于可以检测出放射性同位素的衰变，因此，它们可以示踪被标记的分子以及它们在体内的代谢或转运情况。

^{35}S 标记的甲硫氨酸通常用于示踪蛋白质，因为高比活性的甲硫氨酸易于制备。比活性指的是每单位质量的物质中的放射性同位素活性，它决定于放射性同位素的原子数与非放射性同位素的原子数的比率。此外，^{35}S 还可用于标记核苷酸，如：把 dNTP 中的 α 位磷酸基上的氧原子用^{35}S 取代，^{35}S 标记的核苷三磷酸可用于核苷酸序列分析，也可做其它用途，如：末端标记 DNA 分子等。

^{32}P 常用于标记各种核苷三磷酸。^{32}P 可以是核苷三磷酸的 γ 位磷酸，也可以是 α 位的磷酸，^{32}P 标记的核苷酸是目前在分子生物学技术中用途最广的示踪物之一。

^{14}C 和^{3}H 标记的氨基酸、核苷酸也常用于分子生物学研究。

一、放射性同位素的检测方法

检测掺入的放射性同位素的方法有二种：

（一）放射自显影

在标记了放射性同位素的细胞、组织、切片或其它生物样品上，压上一张 X 光胶片，放射性同位素的电离辐射作用于 X 光胶片上的乳胶，就像可见光一样产生了一个潜在的图象。因此，经过曝光，射线与乳胶作用产生电子，电子使卤化银还原成金属银，这样，便得到了放射性同位素的分布图。在细胞水平上，放射自显影可用来确定生物大分子的合成部位，以及随后在细胞内的运动情况。如：通过掺入^{3}H 标记的脱氧胸腺嘧啶核苷酸，我们得知细胞核是 DNA 合成的主要部位。细胞成分的分离和组织学染色也证实，大部分 DNA 在细胞核中。与此相反，通过掺入标记的尿嘧啶到 RNA 中，我们知道 RNA 是首先在细胞核内合成的，但是在细胞中，大部分 RNA 存在于细胞质中。通过掺入标记的氨基酸我们得以了解蛋白质是在细胞质中合成的。也正是由于电子显微镜放射自显影技术的应用，我们看到了蛋白质分泌的途径。

（二）放射性同位素的定量检测

放射性同位素的衰变使空气电离产生电子，这些电子可用盖革计数器检测出来。因此，可以定量测定标记的样品中含放射性同位素的种类和数量。应用闪烁计数器也是定量测定放射性同位素的一种很好手段。闪烁计数器的工作原理是这样的：放射性同位素的衰变产生了粒子，荧光剂吸收了这些粒子的能量后会发出荧光。因此，把放射性同位素与荧光剂混合，测定荧光的闪烁次数，便可知道放射性物质的含量。

检测方法以及放射性同位素的选择都取决于实验目的，标记的化合物掺入细胞组分后必须具有足够高的比活性，如：^{3}H 标记的核苷酸前体就比 ^{14}C 标记的比活性要高。因而在掺入时间较短或细胞样品较少的情况下，通常选用 ^{3}H 标记 RNA 或 DNA。

在放射自显影的研究中，放射性同位素衰变所产生的粒子的能量对细胞定位的精确性有影响。如：^{32}P 的衰变释放出高能量的硬 β 射线，它们在 X 光胶片上经过显影留下长达 1 mm 的痕迹，比一个单细胞的直径还要长。因此，不适于用作细胞定位。正相反，^{3}H 所释放的 β 射线在 X 光胶片上仅产生 0.47 μm 的黑影，因此，在细胞内定位可达到 0.5～1.0 μm 的精确。

二、脉冲标记实验

脉冲标记实验也就是在要标记的样品中加入放射性标记的化合物，经过短时间后，将放射性标记的化合物取走，同时加入过量的未标记的化合物。脉冲标记实验通常可以示踪标记的化合物分子在不同的细胞器以及生物代谢途径的不同分子之间的移动情况。

在标记的化合物分子，如：氨基酸、核苷酸或磷酸等被掺入以前，它们首先进入到细胞内的构件分子库中。这是一个由可以在细胞质和细胞核中自由扩散的小分子所组成的库，这些小分子中包括氨基酸、核苷酸等，它们有的可自由扩散通过细胞内原生质膜而进出不同的细胞器，有的则不行。

随着细胞生长条件的不同，组成细胞构件分子库的各组分的含量也发生变化。同样，细胞吸收、利用、分泌不同的构件分子的速度也有变化。

氨基酸可以在构件分子库和介质间进行迅速交换。如果把正在生长的细胞放入到一个含有放射性标记的氨基酸的介质中，那么只需 5 min，细胞内的构件分子库的放射性比活性达到最高值，蛋白质中的放射性积累稍慢些。因为，标记的氨基酸必须首先掺入到氨基酸库中，但是如果把标记物移走，加入未标记的氨基酸，几分钟内，标记的氨基酸就停止掺入蛋白质了。这是因为细胞内的氨基酸与介质中的氨基酸已迅速达到平衡，因此能观察到非常明显的脉冲标记效果（图 14-1 a）。

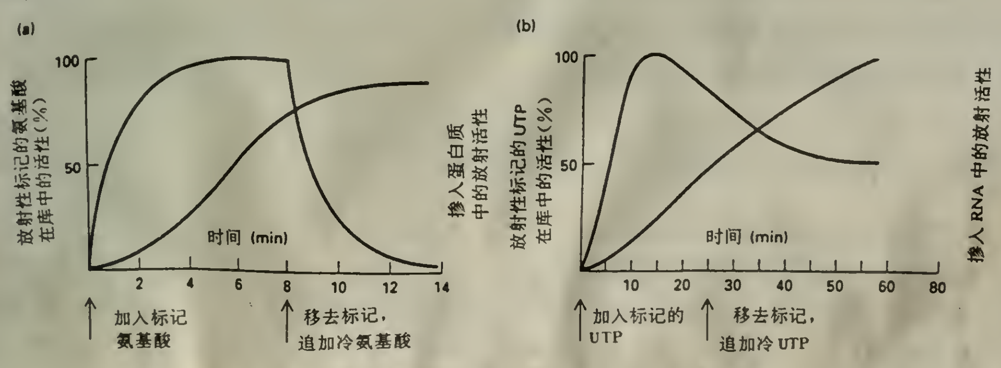


图 14-1 氨基酸与核糖核苷酸的脉冲标记结果比较

然而，核糖核苷酸和脱氧核糖核苷酸进入细胞后不久就被磷酸化，磷酸化后的核苷酸一般不能离开细胞，不能与细胞外的标记核苷酸达成平衡，因此，虽然脉冲标记的尿嘧啶核糖核苷酸可在 10 min 内掺入到库中，但是当标记的尿嘧啶核苷酸移走后，追加的过量的未标记尿嘧啶核糖核苷酸不能有效地降低库中的标记物的含量。这些标记的尿嘧啶核糖核苷酸最终要被掺入到 RNA 中，该过程约需几小时，因此，我们看不到那种明显的脉冲标记效果（图 14-1 b）。

与此相反，脱氧核糖核苷酸的脉冲标记效果很明显，这是因为细胞的库中的脱氧核糖核苷酸含量在 DNA 合成时仅够几分钟用。追加未标记的脱氧核糖核苷酸能有效地降低库中相应的标记物的水平。

因此，在设计脉冲标记实验和解释实验结果时，应该充分考虑到各种前体小分子在细胞的库中所具有的特性。

三、放射性同位素技术在分子生物学中的应用

放射性同位素除了能够标记生物大分子，定量测定它们在细胞器、细胞和组织中的含量以及它们在细胞内的分布外，还能够用于研究生物大分子的合成，下面仅举一例予以说明。

当放射性标记的前体小分子进入细胞后，它们即被掺入到正在合成的大分子中，如果把放射性同位素标记的氨基酸加到介质中，它们即被掺入到正在合成的多肽链中。随着掺入时间的延长，越来越多的已完成的多肽链被标记上。在不同的时间间隔取样，当完整的多肽链中所含放射性同位素的量与还连在模板上的多肽链上的同位素量相等时，所需的时间，就是合成一条特异多肽链的时间，因为当标记开始时，只有正要起始合成的多肽链从头开始被标记，其余的正在延伸的多肽链中大小不一的部分被标记上，当整条完整的多肽链从 N 端到 C 端都被标记上时，从开始标记起到现在的时间，便是合成多肽链所需的时间，也是完成的与未完成的多肽链中含有相等量的同位素时的时间(图 14-2)。

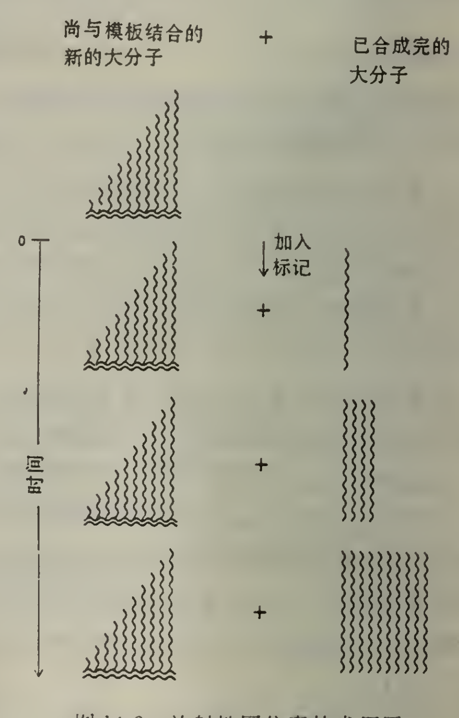


图 14-2 放射性同位素技术用于测定大分子的合成时间

第三节 核酸的提取和纯化

核酸可分为两大类：脱氧核糖核酸，即 DNA；核糖核酸，即 RNA。真核细胞的 DNA 主要存在于细胞核中，与蛋白质相结合构成大小不一的染色体。核外也有少量 DNA，如线粒体 DNA 和叶绿体 DNA 等。DNA 是遗传信息的主要贮藏者和携带者。RNA 主要存在于细胞质中，真核细胞的 RNA 分子的合成发生在细胞核，合成的 RNA 前体分子在核内加工成熟，并穿过核膜运至细胞质中用于指导或帮助蛋白质的合成。RNA 是遗传信息的传递者。因此，要进行分子生物学研究，首要任务就是从组织或细胞中提取核酸。

一、核酸的提取

(一)DNA 的提取

DNA 的提取其最根本的要求是保持核酸的完整性。一般情况下，随着生物由低级到高级

的进化,DNA 的分子量也由小到大递增。如病毒的 DNA 分子量为几至几十千碱基对,细菌的为几千千碱基对,而高等动植物的则达到几百万千碱基对。基因组文库的构建或基因组 DNA 的 Southern 分子杂交等都需要分子量大小至少在 100～200 kb 范围的 DNA,而在提取 DNA 的过程中有许多因素能导致 DNA 降解成小片段:①物理因素降解。因为 DNA 分子量较大,机械张力或高温很容易使 DNA 分子发生断裂。因此,在实际操作时应尽可能轻缓,尽量避免过多的溶液转移,剧烈的振荡等,以减少机械张力对 DNA 的损伤,同时也应避免过高的温度;②细胞内源 DNA 酶的作用。细胞内常存在活性很高的 DNA 酶,细胞破碎后,DNA 酶便可与 DNA 接触并使之降解。为了避免和钝化 DNA 酶的作用,在溶液中常加入 EDTA,SDS 以及蛋白酶等。EDTA 具有螯合 Ca^{2+} 和 Mg^{2+} 的作用,而 Ca^{2+} 和 Mg^{2+} 是 DNA 酶的辅因子。SDS 和蛋白酶则分别具有使蛋白质变性和降解的作用;③化学因素也会降解 DNA。如过酸的条件下,由于 DNA 脱嘌呤而导致 DNA 的不稳定,极易在碱基脱落的地方发生断裂,因此,在 DNA 的提取过程中,应避免使用过酸的条件。

DNA 在体内通常都与蛋白质相结合,蛋白质对 DNA 制品的污染常常影响到以后的 DNA 操作过程,因此,需要把蛋白质除去。一般采用苯酚-氯仿抽提的方法。苯酚、氯仿对蛋白质有极强的变性作用,而对 DNA 无影响。经苯酚-氯仿抽提后,蛋白质变性而被离心沉降到酚相与水相的界面,DNA 则留在水相,这一方法对于去除核酸(无论是 DNA 或 RNA)中的大量的蛋白质杂质行之有效,因此也是一个基本方法。少量的或与 DNA 紧密结合的蛋白质杂质可用蛋白酶予以去除。

DNA 制品中也会有 RNA 杂质,但 RNA 极易降解。况且,少量的 RNA 对 DNA 的操作无大影响,必要时可加入无 DNA 酶的 RNA 酶以去除 RNA 的污染。

DNA 的提取没有统一的方法,但它们的第一步都是将细胞破碎,然后去除蛋白质以及多糖等的污染。细胞器的 DNA 提取,一般采取先提纯细胞器,再溶解细胞器膜,待释放出细胞器 DNA 后,再进一步提取纯化之。利用差速离心通常能达到纯化细胞器的目的。如:线粒体或叶绿体的提取。细菌质粒 DNA 的提取是根据环状质粒 DNA 分子具有分子量小,易于复性的特点进行的。在热或碱性条件下 DNA 分子双链解开,若此时将溶液置于复性条件,由于变性的质粒分子能在较短时间内复性而染色体 DNA 不行。经过离心,上清液是质粒 DNA 分子,而沉淀是变性的染色体 DNA 和蛋白质杂质,以此使质粒 DNA 与染色体 DNA 分离。噬菌体 DNA 的提取则通常先沉淀噬菌体分子,经 CsCl 密度梯度离心后,噬菌体得以纯化,再用苯酚-氯仿抽提去除噬菌体蛋白,即得到噬菌体 DNA。

(二)mRNA 的提取

细胞内主要有 3 种 RNA。mRNA 只占 RNA 总量的 5%左右。其分子量大小不一,由几百至几千个核苷酸组成。mRNA 主要存在于细胞质中,大部分 mRNA 均与蛋白质结合在一起形成核蛋白体。mRNA 的提取对于进一步研究基因的结构及其表达调控是必要的。原核生物的 mRNA 半衰期较短,提取困难。而真核生物的 mRNA 较稳定,相对来说比较容易提取。

要成功的提取 mRNA,关键在于尽可能完全抑制或去除 RNA 酶的活性。RNA 酶是非常稳定的酶,它由一条多肽链组成,变性后容易复性,因此很耐热,煮沸后还能保持大部分活性。RNA 酶无需任何辅助因子,在较宽的 pH 范围内都有活性。操作者的手、唾液等含有较丰富的 RNA 酶。因此,mRNA 的提取方法,也是根据尽量减小 mRNA 的降解,把 RNA 酶的活性降低到最低限度的原则而设计的。

提取mRNA一般有二条途径。其一是先提取多聚核糖体，再将蛋白质与mRNA分开。利用抗原抗体的反应，还可以将含量极微的、特异的mRNA提取出来。因为没有合成完的蛋白质还停留在多聚核糖体上。这些新生肽链能与完整蛋白质的抗体发生抗原抗体的反应，因此，可以选择性沉淀特异的mRNA。其二是提取细胞总RNA，利用寡聚脱氧胸腺嘧啶核苷酸纤维素柱层析[oligo(dT)-cellulose]，把mRNA与其它的RNA，如tRNA，rRNA分开，因为几乎所有的mRNA 3′端都具有poly A的序列，而tRNA和rRNA上都没有这样的结构。进行oligo(dT)-纤维素柱层析时，只有3′端有poly A的mRNA才与oligo(dT)结合，从而达到与其它RNA分离的目的。

在所有RNA的操作中，凡接触RNA的器皿必须严格消毒，溶液中必须加入RNA酶的抑制剂，如RNasin。氧钒基-核苷复合物、异硫氰酸胍等，所用溶液和水一般先用焦碳酸二乙酯处理，再经高温灭菌。操作者必须戴塑料手套。

二、核酸的纯化

核酸由4种不同的核苷酸组成。它们之间在化学性质上差异不大。因此，根据核酸的长度大小的不同而进行分离纯化就成了最常用的方法。

（一）密度梯度离心

超速离心是生物化学中常用的分离大分子的手段。超速离心又可分为两类：一类是根据样品的质量来进行分离，称之为速度区带离心。样品铺在梯度介质的顶部，在离心过程中以不同的速度往下沉降。沉降的速度与样品的质量、密度、离心力的大小、介质的密度以及样品与介质的摩擦力大小有关。离心完成时，不同的样品在离心管中位于不同的区带上。但是，速度区带离心时间不能太长，否则，样品会沉降至离心管底部。60 000 r/min的速度可沉降分子量为10 kD的分子。蔗糖常选作梯度介质，因此，这一类型的超离心又被称为蔗糖密度梯度离心。另一类超速离心则根据样品的密度大小不同而进行分离，称之为密度梯度离心，平衡密度梯度离心是分子生物学研究中最常用的一种。

密度梯度可在离心开始以前做好，也可在离心时依靠离心力自己形成密度梯度。后者就是平衡密度梯度离心技术。在上面所说的两种情形中，介质的密度梯度既要比样品的密度大，又要比样品的密度小。离心时，样品或往上或往下移动，当介质的密度与样品自身的密度相等时，样品停止移动，并停留在该介质密度的位置上。甚至在很大的离心力作用下，样品也不会沉降到比它自身密度大的位置(图14-3)。

形成密度梯度的介质，最常用的是氯化铯(CsCl)的水溶液。由于铯离子(Cs^+)的分子量较小，但其密度大于溶剂分子——水的密度，在离心场作用下，CsCl分子在溶液中重新分布而形成了密度梯度。离心管底部的溶液较之上部溶液含有更多的CsCl分子，一般情况下，顶部与底部溶液之间的密度差可以达到0.02 g/ml。因此，该技术可以方便地分离密度差仅为0.02 g/ml的分子。在CsCl溶液中，蛋白质，DNA和RNA的密度分别为1.3，1.6～1.7和1.75～1.85 g/ml，因此它们能被容易地分离开。这里所给出的密度值比相应的大分子在活细胞中所具有的密度值要高。这是因为CsCl溶液中的离子可结合到蛋白质或核酸上去。如果没有结合的离子，那么所有这些大分子都具有几乎相等的密度值，即1.25～1.3 g/ml。在CsCl溶液中，Cs^+主要结合在DNA的磷酸基上，而对RNA，除了结合在磷酸基上外，还结合在核糖的羟基上，因而RNA的密度比DNA大。蛋白质的平均电荷比核酸少，结合的Cs^+也少。

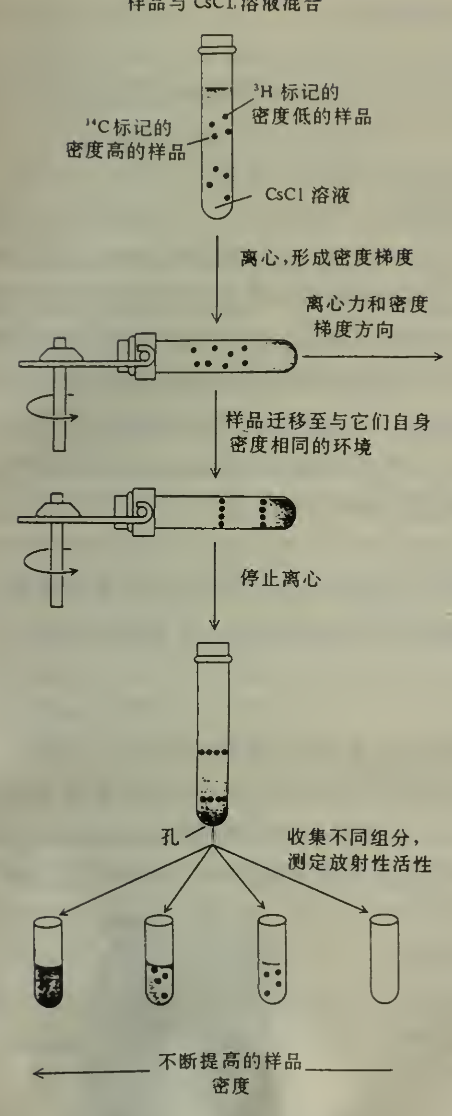


图 14-3 平衡密度梯度离心

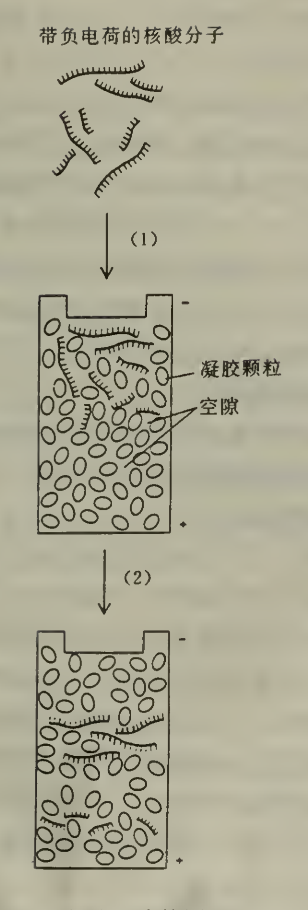


图 14-4 凝胶电泳的原理

(1)混合物置于琼脂糖或聚丙烯酰胺凝胶中,施加电场;

(2)分子以它们的链长成反比的速度通过凝胶间的空隙。

如果蛋白质或核酸中的^{12}C或^{14}N被^{13}C或^{15}N所代替,那么这些蛋白质或核酸就能从含有正常分子的混合物中分离出来,如在蛋白质中,N 的含量占 14%,^{15}N的密度比^{14}N大 15/14 倍,若蛋白质中的^{14}N完全被^{15}N所取代,它的密度将比正常的蛋白质高 1%,这已足以使正常的与被取代的蛋白质得到完全分离。因此,如果细胞在含有重的氨基酸或核苷酸前体的培养基中生长,就有可能分离重同位素加入前后细胞所制造的大分子。

DNA 的密度与其G-C 对的含量成正比。这是因为G-C 对中有 3 个氢键的缘故,G-C 对与 DNA 密度 ρ 的关系可由下列公式表示:

$$\rho=0.098(G\%-C\%)+1.680$$

双链 DNA 中,G 和 A 分别与C 和 T 严格配对,因此知道 G-C 对的含量后,便可推算出

A-T 对的含量，故可利用超速离心对 DNA 的碱基组成进行测定。

（二）电泳

电泳能够分离混合物中大小不同的分子。在电场作用下，溶液中的带电粒子的移动速度决定于它们的电荷-质量比。

核酸在溶液中由于它们有磷酸基而带负电荷，在电场中核酸向正极移动。但是，由于每一个核苷酸都具有几乎相同的电荷-质量比，长度不同的核酸分子也具有几乎一致的电荷-质量比。同样的，蛋白质分子的形状、质量可以不同，电荷-质量比却是相近的，因此，如果只在溶液中进行核酸或蛋白质的电泳，那就不会使大小不同的分子得到分离。但是，如果把蛋白质或核酸的电泳改在凝胶中而不是在溶液中进行，这样的分离就能实现。这一种电泳技术称之为凝胶电泳（图 14-4）。常用的凝胶有聚丙烯酰胺凝胶和琼脂糖凝胶，凝胶的孔径大小限制了分子移动的速度。具有相同荷质比但长度不同的分子移动的速度不同，长的分子移动速度慢。即便是由 10 000～20 000 个核苷酸组成的核酸分子，相差几个百分点也能分离开。长度不超过 500 个核苷酸的分子，如果相差仅一个核苷酸，也能在电泳中被分离开，这也就是 DNA 序列分析得以实现的基础。

琼脂糖凝胶电泳是 DNA 电泳的常规方法。它可用来测定 DNA 的分子量，分离经限制性内切酶水解后的 DNA 片段，进一步纯化 DNA 等。琼脂糖凝胶的孔径大小主要取决于凝胶中琼脂糖的浓度。

RNA 也可用琼脂糖凝胶电泳进行分离鉴定。由于 RNA 分子容易形成二级结构，从而影响 RNA 分子的迁移率。因此，RNA 的电泳中常加入变性剂，如：甲酰胺、尿素、甲醛和乙二醛等。

聚丙烯酰胺凝胶电泳常用于蛋白质及小分子量核酸的分离。其孔径大小主要取决于丙烯酰胺与甲叉双丙烯酰胺（交联剂）的浓度。DNA 的序列分析就采用聚丙烯酰胺凝胶电泳。蛋白质的 SDS（十二烷基硫酸钠，去污剂）聚丙烯酰胺凝胶电泳可用于测定蛋白质的分子量，中性 pH 下，SDS 带负电，蛋白质的每个氨基酸残基结合一个分子的 SDS，相邻的 SDS 分子由于带有相同的负电荷而相互排斥，使结合了 SDS 分子的蛋白质形成棒状构型并且具有相同的荷质比，此时的蛋白质处于变性状态。它们在凝胶中的迁移速度决定于它们的长度。相差 1% 的蛋白质也能被分离。

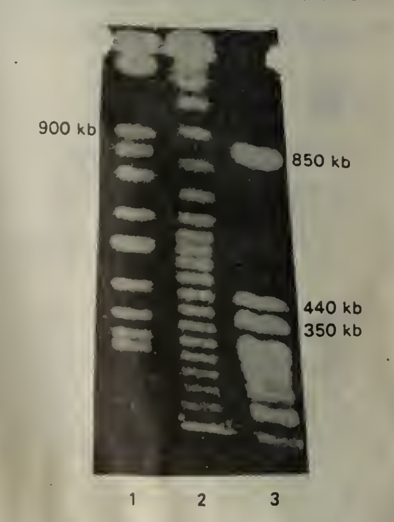


图 14-5 脉冲场凝胶电泳图谱

1. 酵母染色体 DNA，其中每一条带代表一条染色体；2. λ 噬菌体 DNA 的连环体，最小单位分子量为48.5 kb；3. *Not*I 消化的 *E. coli* 染色体。

脉冲场凝胶电泳技术是近年来发展起来的分离大分子量 DNA 的有效手段。线状的双链 DNA 分子超过一定的长度后，它们在琼脂糖凝胶电泳中以相同的速度泳动，这是因为琼脂糖凝胶的孔径太小。降低琼脂糖的浓度虽然可以有限地增大凝胶的孔径，但是这样的低浓度（0.1%～0.2%）凝胶不易操作，而且，超过 750 kb 的 DNA 分子不能得到有效分离。而脉冲场电泳技术可用于分离分子量范围在 1～10 Mb 的线状双链 DNA 分子。当 DNA 分子在电泳中朝一个方向移动时，DNA 分子沿着电场的方向伸展开，这时，切断电流，DNA 则开始回复松弛的无规卷曲状，松弛所需的时间取决于 DNA 分子的长度。分子越长，回复所需的时间也越长。短时间后，一个新的电场加在与原来电场成一定角

度的方向，长的 DNA 分子因还未回复到松弛状态，需要更多的时间才能重新开始朝新的电场方向移动。这样，在这两个方向上反复施加电场，就能使大分子量的 DNA 分子得到分离（图 14-5）。因此，这一技术在染色体的分析中是不可替代的。因为最小的酵母染色体也有 5×10^5 bp，最大的动物或植物的染色体则有 3×10^8 bp。

三、限制性内切酶的应用

要进一步分析提取纯化的 DNA，首先必须依赖于限制性内切酶的应用。利用限制性内切酶，我们可以把较大的 DNA 分子重复地切割成分子量符合要求的大小片段。也只有利用限制性内切酶，我们才能构建出 DNA 的物理图谱，了解限制片段之间的排列顺序，也才能够进行 DNA 序列分析和目的基因的克隆。

从第三章第七节里，我们认识了限制性内切酶。目前已经发现并提纯的限制性内切酶已达 100 多种。正确选择使用限制性内切酶，主要依赖于我们的实验目的。将要产生的 DNA 限制性片段的分子量大小通常是首要考虑的因素。几百碱基对长短的 DNA 片段适于进行精细的物理图谱以及 DNA 序列分析等。因此常选用切点较多的识别位点为 4 个核苷酸序列的酶。这样的酶也适于进行 DNA 的部分酶解。这在建立基因组文库时是常用的方法。

识别位点为 6 个核苷酸序列的限制性内切酶通常产生 1～10 kb 长短的限制性片段，适于进行较长的 DNA 分子的物理图谱分析和包括内含子、结构基因以及调控序列在内的完整基因的克隆。

识别序列为 8 个核苷酸的限制性内切酶较少，但是对于构建更大的 DNA 分子的物理图谱非常有用。它们产生的限制性片段可以在脉冲场电泳中得到分离。

选择限制性内切酶时另一个需要考虑的因素是产生相同的末端以便于连接和克隆。DNA 连接酶能够将具有相同粘性末端或都是平端的两个 DNA 末端连接起来。在 DNA 克隆时，需要把 DNA 的限制性片段与载体 DNA 分子相连接。因此，必须选用产生相同末端的限制性内切酶分别消化所要克隆的 DNA 分子和载体 DNA 分子。在此我们看到，正是由于限制性内切酶和 DNA 连接酶的应用，使人们能够在体外进行 DNA 的切割和连接，为重组 DNA 技术的出现奠定了重要基础，具有划时代的意义。

限制性内切酶相对来说是比较稳定的酶，但是不正确地使用常常也会出现问题，影响酶活性的主要因素有：(a)DNA 底物的纯度。常见的污染物为蛋白质、乙醇、酚、氯仿、过高的EDTA 等；(b)DNA 分子的结构。包括 DNA 分子的超螺旋、甲基化等；(c)反应条件。如：反应温度、缓冲液组成等。

第四节　核酸的分子杂交和 DNA 的人工合成

一、核酸的分子杂交

核酸的分子杂交技术是目前分子生物学中最广泛应用的技术之一。它可以鉴定核酸分子之间的同源性。这一技术是分子克隆技术的重要组成部分，在基因的表达调控和物种的亲缘关系研究中也发挥重要作用。从 DNA 和 RNA 的结构中我们得知：DNA 为双链双螺旋分子，其中一条链与另一条链的核苷酸序列互补，RNA 分子若其顺序与 DNA 分子有互补关系，也能

与DNA形成互补的双链结构。因此，两条来源不同的，具有核苷酸互补关系，也即：具有一定同源性的DNA单链分子或DNA单链分子与RNA分子，在去掉变性条件后互补的区段能够退火复性形成双链DNA分子或DNA/RNA异质双链分子。我们把这一过程称为核酸的分子杂交。核酸的分子杂交有不同的形式和方法。下面我们介绍常用的几种。

(一)Southern 分子杂交技术

以它的发明者命名的Southern分子杂交技术已成为基因分离和鉴定的必不可少的手段，也广泛用于遗传多态性分析。

DNA样品用一种或几种限制性内切酶消化后，DNA片段按照分子量的大小在电泳凝胶中进行分离。凝胶经过变性处理(通常加入碱性缓冲液)，使其中的DNA双链分子变性为单链，然后利用毛细管作用原理或其它物理方法，将凝胶中的单链DNA分子转移并固定到固相支持物表面上(一般为硝酸纤维素膜或尼龙膜)，此时的DNA片段还保留着原来在凝胶中的分布图式。如果用一段放射性标记的核苷酸序列(称之为探针)与膜上的DNA片段杂交，那么，与探针有同源性的DNA片段在膜上的位置便可通过放射自显影而得以了解。Southern分子杂交技术具有很高的灵敏度，可以检测出相当于人的基因组中只出现一次的DNA序列(约$1/10^6$)。它与绘制DNA物理图谱相结合，可以确定目的基因(探针或同源序列)在图谱中的位置及其限制性酶切位点的分布。

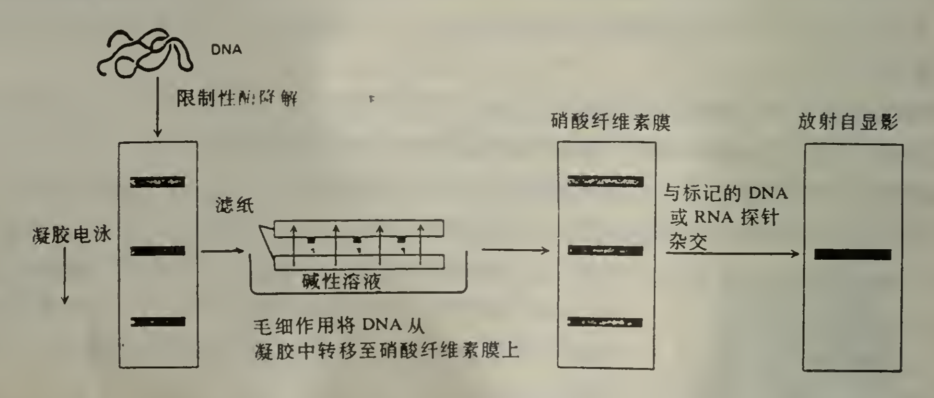


图14-6 Southern分子杂交技术用于检测特异DNA序列的存在

(二)Northern 分子杂交技术

Northern分子杂交技术由于与Southern分子杂交技术类似而得名。它们所不同的是：Northern分子杂交技术用于检测特异的mRNA分子的存在。mRNA样品经过电泳，转移到硝酸纤维素膜上，与放射性标记的DNA探针进行杂交，经过放射自显影，特异的mRNA分子在样品中的存在与否，其长度大小以及量的多寡便可得知。Northern分子杂交技术已广泛用于研究特异的mRNA分子在细胞处于不同条件下发生的量和质的变化或不同组织器官中，基因表达的差异。

另一项类似的技术称为Western蛋白质杂交技术。在这项技术中，经过单向或双向电泳分离的蛋白质被转移到硝酸纤维素滤膜上，放射性或用酶标记的特异抗体即可用来检测相应抗原(蛋白质)的存在。

(三)原位杂交技术

有两种意义的原位杂交技术：一种是指在构建基因组文库或cDNA文库时，长在培养皿

上的菌落或噬菌斑经转移到硝酸纤维素膜上后进行分子杂交。我们将在下一节继续讨论这一技术。另一种原位杂交技术则是指在细胞、组织切片上直接进行分子杂交。通过检测细胞内特异 mRNA 的存在与否，它能够告诉我们，在哪些细胞中有特异基因的表达。要进行原位杂交，首先要进行组织的固定，以免在以后的操作过程中组织的天然形态结构遭到破坏。常用的有甲醛、戊二醛等交联固定剂。以后依次经过脱水、石蜡包埋和切片，组织样品的准备就算完成。以上这些过程与制备显微镜下观察所需的石蜡切片并无太大区别。

原位杂交所用的探针要短，这样才能尽量使探针掺入细胞内。同时，在杂交进行以前常用蛋白酶消化，使细胞的空隙增大，从而提高探针对靶子 mRNA 序列的可接近度，标记探针的同位素常用^{32}S 和^{3}H，^{32}P 因为具有过高的能量导致分辨率降低而往往成为落选者。

原位杂交所需杂交液的用量很小，杂交反应可在盖玻片下进行。与前面所介绍的分子杂交技术不同，原位杂交的放射自显影不是利用 X 光片而是将乳胶涂在切片上，经过曝光、显影、定影来完成。然后，进行组织切片的化学染色，便可观察到切片组织的形态结构以及放射自显影信号在组织、细胞中的位置。原位杂交技术可以检测出一个细胞中只有一个分子的 mRNA。

二、DNA 的人工合成

在核酸的分子杂交技术中所使用的探针可以是来源于其它物种或自身的同源序列，也可以是化学合成的寡聚核苷酸序列，这尤其适用于只知道特定目的 DNA 片段的部分序列时的情况（见第一节介绍）。

化学合成寡聚核苷酸序列今天已经能够在机器上自动进行。它采用的是固相磷酸三酯法和亚磷酸三酯法。由于亚磷酸三酯法具有反应速度快，合成效率高和副反应极少的优点，已经在机器自动合成中被广泛使用。其基本原理（图 14-7）是将所要合成的寡聚核苷酸链的 3′-末端先以 3′-OH 与一个不溶性载体，如多孔玻璃珠（CPG）连接，然后依次从 3′→5′的方向将核苷酸单体逐个加上去。所使用的核苷酸单体的活性官能团都是经过保护的。其中 5′-OH 用 4，4′-二对甲氧基三苯甲基（DMT）保护，3′端的二异丙基亚磷酰胺上磷酸的-OH 用甲基或 β 氰乙基保护，碱基上的活泼基团如氨基也都用不同的保护基予以保护。合成的延伸是这样开始的：

图 14-7　DNA 的化学合成原理

（一）脱三苯甲基保护基

加入 $ZnBr_2$ 或三氯乙酸，脱去 4，4′-二对甲氧基三苯甲基，释放出与载体相连的寡聚核苷酸单体 1 上的 5′-OH。

（二）缩合反应

在弱酸-四唑的存在下，新加入的带保护基的核苷酸单体 2 与单体 1 发生缩合反应，形成 3′，5′-磷酸二酯键，其中磷为三价。

（三）盖帽反应

加入乙酸酐，使极少数尚未参与缩合反应的核苷酸单体 1 中的 5′-OH 乙酰化，从而达到使其封闭，终止其以后参与缩合的资格，以减少发生合成错误序列的机会。

（四）氧化反应

形成的 3′，5′-磷酸二酯键由于磷为 3 价，也称为亚磷酯键。亚磷酯键不稳定，能被酸或碱解离，故加入碘将其氧化，成为比较稳定的 3′，5′-磷酸二酯键，其中的磷为 5 价。

如此经过多次循环，直到合成出具有所需序列长度的寡聚核苷酸片段后，将其从固相载体中释放下来，再脱去其余的所有保护基，就得到了所需的寡聚核苷酸。凝胶电泳法或高压液相色谱法可用于它们的进一步纯化。

化学合成寡聚核苷酸现在已经在分子生物学研究中得到了广泛应用。它能够：①作为核酸分子杂交的探针；②作为酶促 DNA 合成的引物，用于 DNA 序列分析，聚合酶链式反应（PCR）等；③作为基因合成的元件。对于一些比较小的基因或天然来源不易得到完整基因的或对天然基因进行部分改造时，可以利用人工合成基因的办法获取目的基因。化学合成的寡聚核苷酸片段则担当起人工基因合成的元件。常用的组装基因的办法有两种，即：小片段的互补粘接法和大片段的部分互补酶促延长法（图 14-8）。合成出的双链 DNA 作为整个基因合成的一个段块，与其它的段块连接成更长的双链DNA，最后组装成一个完整的基因；④作为DNA末端改造

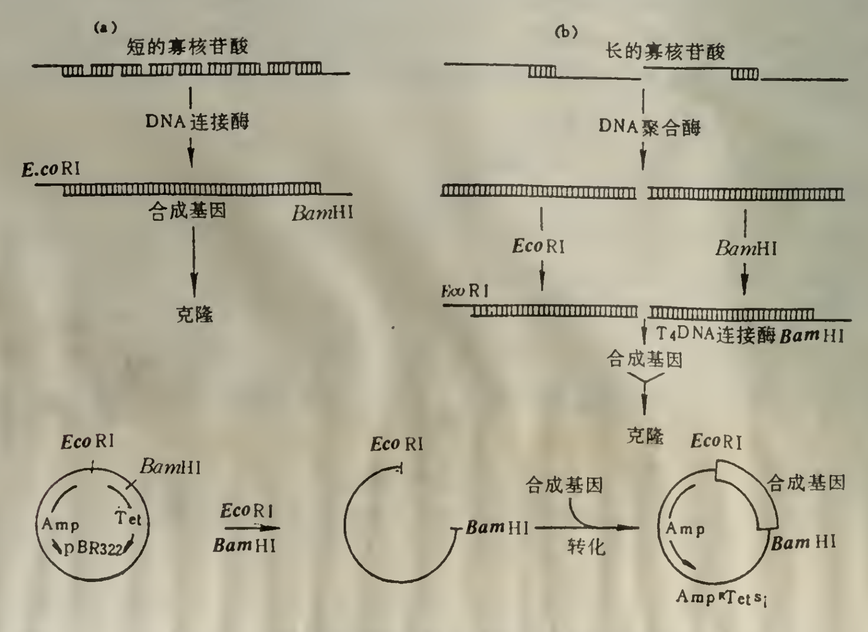


图 14-8　小片段的互补粘接（a）和大片段的酶促延长（b）

中的接头。基因克隆过程中，往往需要在 DNA 末端引入新的限制性内切酶位点。这些新的位点可以通过化学合成的二条互补的寡聚核苷酸链经过退火后形成。再经与所要克隆的目的 DNA 片段连接，便形成了具有新的限制性内切酶位点的末端。

第五节　目的基因的分离

目的基因的分离无疑是分子生物学的重要内容。要进行基因的结构和功能的研究，首先应该得到纯化的、单一的目的基因。利用限制性内切酶可以直接分离得到那些定位在质粒或病毒上的基因。但是，如果要分离得到人的基因组中的某一单一的 DNA 片段则要困难得多。因为人的染色体DNA 总共长约有 3×10^9 bp。假如限制性内切酶每隔 3 000 bp 切割一次，将产生上百万的不同的 DNA 片段，可以设想，利用电泳及其它已经介绍过的技术，要想纯化出其中的某一 DNA 片段几乎是不可能的，重组 DNA 技术却能够克服这种从较大的基因组中分离纯化单一的目的基因的困难。

利用重组 DNA 技术分离目的基因，我们也称之为基因克隆。这里的“克隆”一词是动词，是指从单一祖先产生同一的 DNA 分子群体或细胞群体的过程。而“克隆”用作名词时，则是指从一个共同祖先无性繁殖下来的一群遗传上同一的 DNA 分子、细胞或个体所组成的特殊的生命群体。因为在基因克隆的过程中，含有目的基因的外源 DNA 经过限制性内切酶消化或其它手段形成 DNA 片段后，在体外与能在细胞中独立自主复制的载体分子连接，形成的重组 DNA 分子被引入寄主细胞，每一个含有重组 DNA 分子的寄主细胞增殖后成为一个克隆，其中带有一段同一的外源 DNA 片段，经过筛选，便可以从大量的克隆中得到含有目的基因的 DNA 片段的克隆，同时，插入的 DNA 片段也得到扩增。

由大量的含有基因组 DNA（即某一生物的全部 DNA 顺序）的不同的 DNA 片段的隆克所构成的群体，我们称之为基因文库。一个完全的基因文库，应该能够保证从中筛选到目的基因。1976 年，L. Clark 和 J. Carbon 提出了一个完全的基因文库所需克隆数的计算公式：

$$N=\ln(1-p)/\ln(1-f)$$

在上述公式中：N 是一个完全基因文库所应包含的重组体克隆数；p 为所期望的目的基因在基因文库中出现的几率；f 是插入片段的平均大小与基因组 DNA 大小的比值。

从这一公式中，我们可以计算出不同来源的基因组构建基因文库时需要的最少克隆数的理论值。

若这些大量的重组体 DNA 分子所含的外源 DNA 不是基因组 DNA，而是由某一生物的特定器官或特定发育时期细胞内的 mRNA 经反转录形成的 cDNA 分子，它们所构成的重组 DNA 克隆群体，则称为 cDNA 基因文库。

在基因文库和 cDNA 文库中所使用的寄主细胞通常为大肠杆菌 *E. coli* 细胞。最近，酵母的人工染色体系统（YACs）技术已经出现。利用此技术可以方便地克隆非常大的基因组 DNA 片段。

一、cDNA　克　隆

网织红细胞中，含有大量的珠蛋白 mRNA，细胞内合成的蛋白质 90%以上是 α 和 β-珠蛋白。下面，我们以珠蛋白的 cDNA 克隆为例，说明建立 cDNA 文库的原理，见图 14-9。

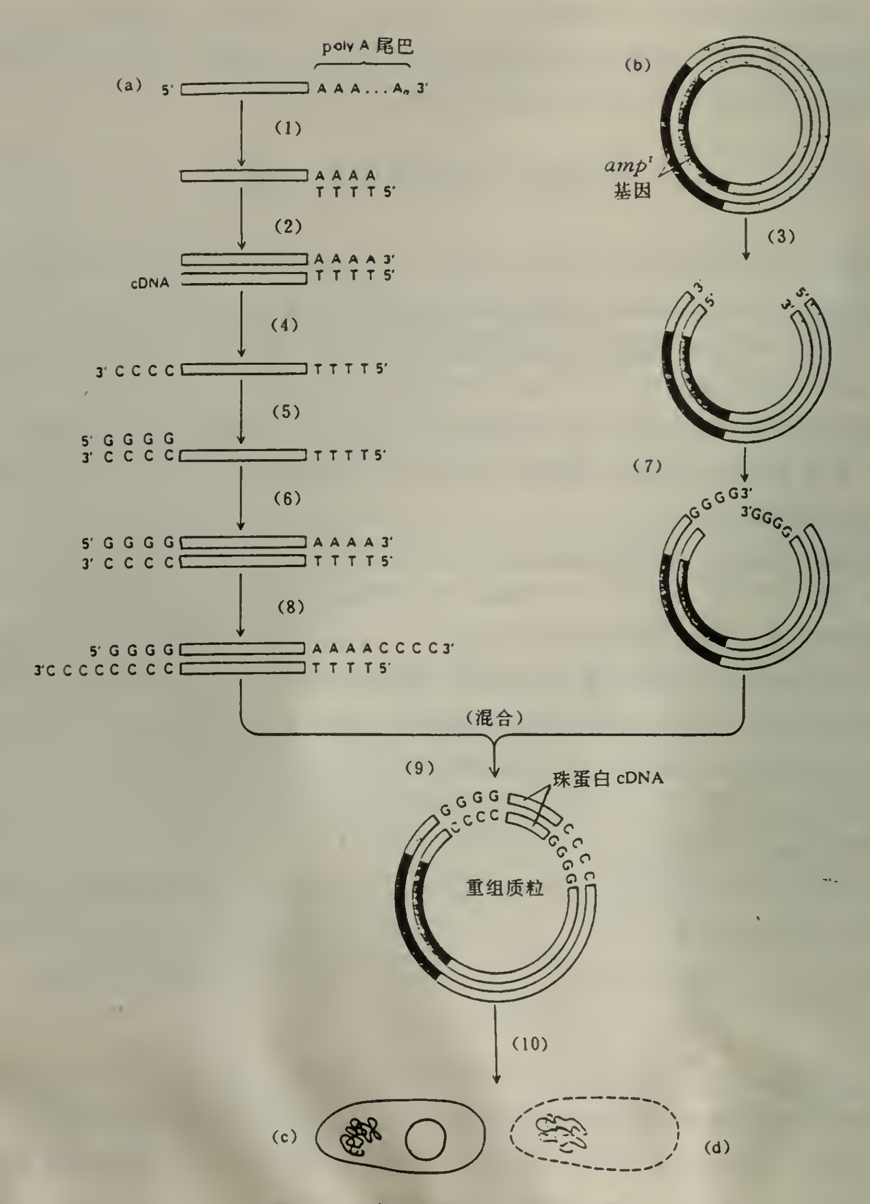


图 14-9 编码珠蛋白的 cDNA 克隆

(a)来自网织红细胞的珠蛋白 mRNA；(b)带有氨苄青霉素抗性基因的质粒；(c)含重组质粒的 *E. coli* 细胞生长；(d)没有 amp^r 的细胞死亡。(1)用 poly T 引物杂交；(2)反转录酶从 mRNA 反转录出 cDNA；(3)限制性内切酶消化；(4)碱解 RNA，加 poly C 尾巴到 cDNA 的 3′端；(5)与 poly G 引物杂交；(6)DNA 多聚酶；(7)加同聚物尾巴到 3′端(与双链 cDNA 尾巴互补)；(8)加同聚物尾巴到 3′端；(9)珠蛋白 cDNA；(10)转染无 amp^r 基因的 *E. coli* 细胞，用氨苄青霉素选择 amp^r 细胞。

反转录酶是合成 cDNA 的关键酶。它是以 mRNA 为模板把脱氧核糖核酸逐个加到引物上，由 5′→3′的方向合成出与 mRNA 链互补的 cDNA 链。在反转录酶作用前，必须将一个短的引物杂交到 mRNA 分子的 3′端，幸运的是，真核生物的 mRNA 分子在其 3′端通常有一段长度约为 50～250 个核苷酸组成的多聚腺嘌呤核苷酸(polyA)结构。因此，多聚脱氧胸腺嘧啶核

苷酸 poly(dT)可以完全满足作为引物的要求。

在 cDNA 链合成完毕后，模板 mRNA 用碱处理降解，DNA 链不受影响。双链的 DNA 可以 cDNA 链为模板合成出来。其中一种技术路线是首先利用末端转移酶(一种无需模板，催化5′脱氧核苷三磷酸以 5′→3′ 为方向，逐个将脱氧核苷酸分子加到线性 DNA 分子的 3′-OH 末端的酶)，在 cDNA 的 3′-末端加上同聚物(poly dC)尾巴。然后以一段多聚脱氧鸟嘌呤核苷酸(poly dG)作为第二条链合成的引物，在 DNA 聚合酶催化下，合成出双链的 DNA 分子。

下一步是把合成好的双链 cDNA 分子与载体分子相连，这里我们选用的载体分子为质粒。质粒是一种能在细菌中独立进行复制的染色体外 DNA 分子，呈双链环状，经过人工改造的质粒在其上通常有许多限制性内切酶的单个切点。选用一种限制性内切酶将质粒分子切断，注意不要破坏质粒的复制起点。然后分别在双链 cDNA 分子和线性化了的质粒 DNA 分子的两个 3′-末端再一次用末端转移酶加上同聚物尾巴。poly(dC)被加到 cDNA 分子的两端，而与它互补的 poly(dG)被加到线性化质粒分子的两端。当 cDNA 分子与质粒分子混合在一起时，它们可通过突出末端的互补作用相连 ，在 DNA 连接酶的作用下，形成环状的重组 DNA 分子。

在质粒分子上，有一个抗药性基因，我们也称之为选择性标记。因为带有这种质粒的细菌对相应的抗生素不敏感。若在培养基中加入相应的抗生素，没有质粒的细菌就不能生长，而带有相应质粒的细菌便会增殖形成菌落。

经过特殊处理的细菌细胞具有吸收外源 DNA 的能力，我们把它们称为感受态细胞。当重组质粒分子与感受态细胞混合，由于感受态细胞的数量远大于重组质粒分子的数量，因此，在培养基上长出的单菌落可以认为是由每一个重组质粒分子导入一个细菌细胞后经过增殖而形成的。在我们现在所说的例子中，应该有 90％的重组质粒分子编码 α 或 β 珠蛋白，由于 mRNA 分子通常并不完全被反转录成 cDNA 分子。因此，非全长的cDNA也会被克隆，要进行鉴定，必须测定所克隆的插入 DNA 片段的顺序。含有目的基因的cDNA克隆可按照（图 14-11）进行筛选。

二、基因文库的建立

建立基因文库的技术原理与 cDNA 文库是相似的，但是，由于构建基因文库的目的是要分离定位于染色体上的基因，而基因组 DNA 的总长度通常要比某一特定组织器官或细胞中的 mRNA 分子的长度的总和大许多倍。根据 Clarke 和 Carbon 的公式不难知道，一个完全的基因文库所需的克隆数要远远超过 cDNA 文库的克隆数。例如，人和鼠的基因组 DNA 的长度约为 $3\sim4\times10^9$ bp，如果要构建它们各自的基因文库，其中插入片段的平均分子量大小为 20 kb，那么必须具有 10^6 以上的克隆，才能保证每一 DNA 片段在文库中出现的几率为 99％。由于插入片段的大小对基因文库的大小有着直接影响，因此，通常选用 λ 噬菌体 DNA 作为载体分子，它所能克隆的插入片段的长度在 20 kb 左右。λ 噬菌体 DNA 的长度约有 50 kb，但是中间部分的 25 kb 可以删去，代之以外源 DNA 片段。经过这样重组的噬菌体完全具备对 *E. coli* 细胞的侵染和在 *E. coli* 细胞内进行复制的能力。

构建基因文库的第一步是从某一物种的任一组织中提取基因组 DNA。动物的生殖细胞或早期胚，以及植物的叶片等都是常用于提取总 DNA 的材料。然后，用限制性内切酶消化总 DNA，如果选用 *Eco*R I ，经过消化后的 DNA 片段都具有粘性末端，当 DNA 片段的平均长度为 20 kb 左右时终止消化。λ 噬菌体 DNA 也用相同的限制性内切酶 *Eco*R I 降解，*Eco*R I 在 λ

噬菌体DNA上有两个识别位点，因此消化后的λDNA产生一条25 kb长短的片段，这就是位于λDNA中部的非必需片段，此外还有两条位于两端的短的DNA片段，也称臂。λDNA的中间部分可通过电泳而与两条臂分开。这时，可将λDNA的两臂与基因组的DNA片段混合，混合的比例为1∶1，由于粘性末端是互补的，相当于正常的噬菌体DNA大小的λDNA分子就会形成，其中包含有一段长度约为20 kb的鼠的基因组DNA，在DNA连接酶的作用下，重组DNA分子连接起来。再从感染的*E. coli*细胞中提取出噬菌体的外壳蛋白，在体外进行噬菌体的组装。只有长度正常的λDNA分子才能被有效的包装，形成具有感染能力的λ噬菌体(图14-10)。含有重组λDNA分子的噬菌体感染*E. coli*细胞，在培养基上形成噬菌斑。通过原位噬菌斑杂交，我们便可确定含有特异基因的噬菌体在培养基上的位置。所用的探针可以是同位素标记的特异mRNA或cDNA分子以及其它物种的同源基因组DNA(图14-11)。

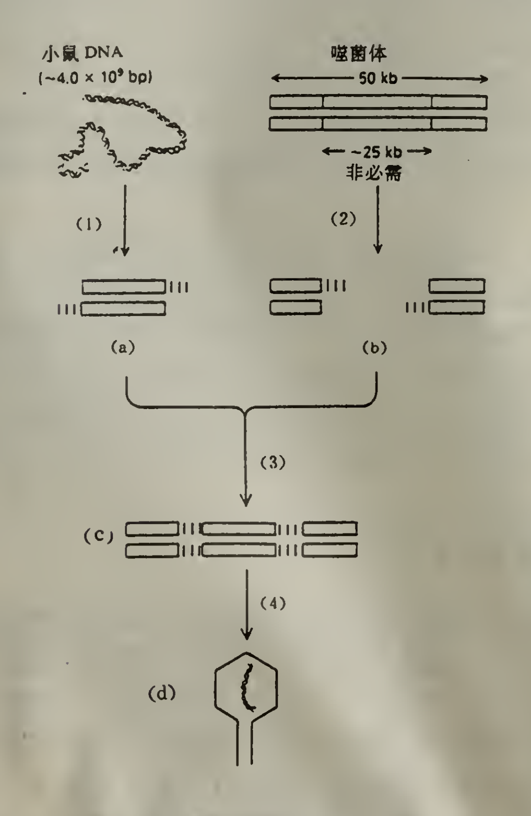


图14-10 建立基因组文库的原理

(1)*Eco*RⅠ酶解成的2×10^5片段；(2)*Eco*RⅠ酶解，去掉中间部分；(3)小鼠DNA与λ双臂混合，粘性末端；(4)用λ蛋白质进行包装只有插入片段为20 kb左右的噬菌体才复制。(a)带粘性末端的20 kb片段；(b)带粘性末端的λ双臂；(c)大小适于包装的重组DNA；(d)含有小鼠基因组DNA的λ噬菌体(为了使文库有90%～95%的可能性包含整个小鼠基因组，需要有10^6这样的噬菌体)。

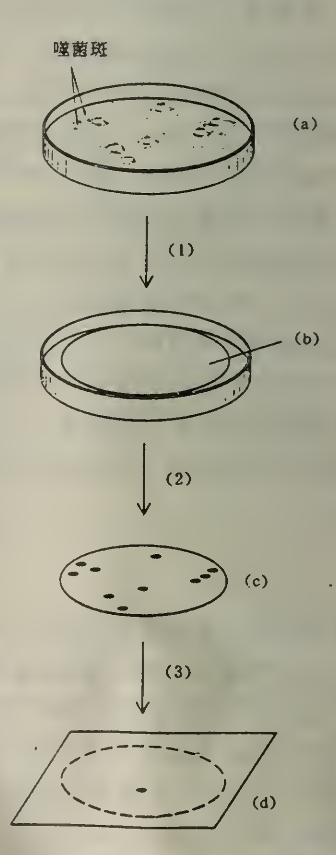


图14-11 利用原位杂交法从基因组文库(或cDNA文库)中筛选得到目的基因

(1)将硝酸纤维素膜置于平板上，以吸附噬菌斑，便于筛选出所需噬菌体；(2)碱性溶液裂解噬菌体，使释放出的DNA变性；(3)与标记的cDNA或mRNA探针杂交，放射自显影。(a)用噬菌体文库(约含2×10^6噬菌体)感染细菌，并铺到平板上以产生噬菌斑；(b)硝酸纤维素膜；(c)单链的噬菌体DNA结合到膜上；(d)与探针序列互补的噬菌体DNA出现阳性信号。

三、特异基因片段的分离

如何才能分离得到含有目的基因的重组DNA 分子，这是基因克隆的中心问题。基因文库或cDNA 文库中的目的基因可利用核酸分子杂交技术，抗体与抗原的特异反应来筛选获得。

不同的生物常具有相同的基因，这些基因的核苷酸序列虽然不尽相同，却享有一定的同源性，因此，如果另一种生物的同一基因已被克隆，那么，它便可以用作探针，筛选得到所需的目的基因。

组织特异性表达的基因可以通过差别筛选法获得。具体的方法是这样的：提取 A 类细胞的 mRNA，构建cDNA 文库。另外，标记从mRNA 反转录形成的单链cDNA 分子，以此由不同的单链 cDNA 分子组成的探针与所构建的 cDNA 文库进行分子杂交。再提取 B 类细胞的mRNA，同样标记反转录后形成的单链cDNA 分子，也以 B 类细胞的单链 cDNA 为探针，与 A 类细胞的cDNA 文库进行分子杂交。如果 A 类细胞中有相对于 B 类细胞来说是特异性表达的基因，那么，该基因应该只与 A 类细胞的单链 cDNA 探针杂交，而不与 B 类细胞的探针杂交。与两类单链cDNA 探针都杂交的克隆则是在 A 类和 B 类细胞中都表达的基因。而与两类探针都不杂交的克隆则可能是由于其mRNA 丰度较低导致无法检测出其存在与否。这一种筛选目的基因的方法也称为正负筛选法。同样原理，我们也可筛选得到不同的发育时期、不同环境条件下特异性表达的基因。

合成的寡聚核苷酸序列也可作为探针，这尤其适用于只了解目的基因的产物——蛋白质的部分序列时的情况。因遗传密码具有简并性，从蛋白质推导出的核苷酸序列与目的基因的相应序列可能并不一致，但这并不影响合成的寡聚核苷酸探针成为一种有效的筛选手段。

如果目的基因所编码的蛋白质易于提取，利用抗体对抗原的特异性识别进行筛选也是一种很好的方法。其中的抗原就是提取得到的蛋白质，制备出该抗原的抗血清后，构建一个表达的 cDNA 文库。由于选用了表达载体分子，插入的基因可融合到一个大肠杆菌基因的后面，因此，可在大肠杆菌中诱导表达，当抗体分子识别出表达了相应抗原的克隆时也就是找到了相应的含有目的基因的重组 DNA 分子。

如果打算克隆一个未知的真核生物基因，也就是说，既没有目的基因的核酸探针，也没有基因所编码的蛋白质，但是该基因与某一表型有关。在这种情况下，通常采用的方法是引入一个示踪核苷酸序列到真核生物细胞内，该序列可插入到染色体上的某一位置，使所插入的位置上的基因结构发生改变，基因的失活引起了可观察到的表型的变化，选择表型变化了的细胞，构建其基因文库，以示踪核苷酸序列作为探针，筛选出含有示踪核苷酸序列的插入片段，在此片段中，位于示踪序列两边的核苷酸序列就应该是与该表型有关的基因的序列。常用的示踪核酸有病毒分子如反转录病毒或转座子等，它们进入真核细胞后都具有整合到染色体上的能力。

分离目的基因，离不开构建基因文库或cDNA 文库。如何决定选用哪一种文库呢？首先，这取决于我们将采用的筛选方法。例如：若采用核酸探针的筛选方法，则两种文库都适用。如果采用抗体-抗原的特异性反应作为筛选的手段，则应选用构建cDNA 文库的方法。其次，要考虑目的基因在这两种文库中的丰度。一般说来，cDNA 文库较小，一个完全的cDNA 文库所含的克隆数要比一个完全的基因文库所含的克隆数少得多，它们之间相差可达 10～100 倍，如果能恰当地选择mRNA 的来源，从cDNA 文库中筛选得到目的基因的机率就大些。但是有些基因组简单的生物如：酵母、盘基网柄菌等，目的基因在基因文库中的丰度也与它在cDNA 文库中的丰度相当。此

外，从 cDNA 文库中分离得到的基因可在大肠杆菌中表达，因为 mRNA 经过加工修饰后，切掉了内含子。而表达的产物有时对于研究者来说是重要的。从基因文库中得到的目的基因包含了完整的基因结构信息如：5′-上游区的调控序列，外显子、内含子等。因此，要研究基因的完整结构，如确定内含子在基因中的精确位置，则具有来自这两种文库的目的基因是必不可少的。

四、酵母人工染色体(YAC)克隆系统

利用酵母载体产生的酵母人工染色体(YAC)克隆系统可以克隆和分离极大的 DNA 片段(达数百千碱基对)，从而代替 λ 噬菌体载体和粘粒载体构建文库。酵母人工染色体系统中必须包括下列序列(图 14-12)：

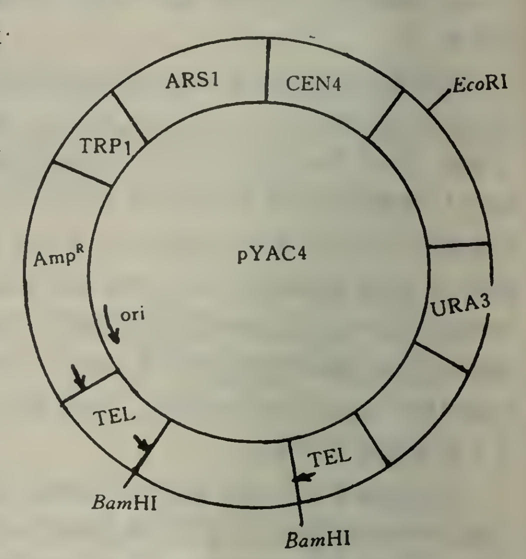


图 14-12 **酵母人工染色体系统**

(一)着丝粒(CEN)

已克隆的真核着丝粒，每一个着丝粒都有一个数字，对应于原先所在的染色体。着丝粒使染色体在分裂过程中能正确分配到子细胞中。

(二)两个端粒(TEL)

它是染色体末端复制所必需的，而且可防止染色体被核酸外切酶逐步降解。

(三)复制起始点(ARS1)

它可使 DNA 在酵母细胞中自主复制。

(四)选择标记

利用它可在酵母细胞中鉴别筛选转化子(如 TRP1 和 URA3)。

(五)原核序列

如 *E. coli* 复制启始位点和 *amp*^{r} 基因。

(六)限制酶切点

便于插入外源 DNA。

利用 YAC 系统克隆的主要过程是，用切割位点较少的限制酶部分酶切基因组 DNA(如 *Eco*RⅠ)用脉冲凝胶电泳分离，纯化出大小为 200～500 kb 的片段。YAC4 载体用 *Eco*RⅠ 和 *Bam*HI 切割，去掉两个 *Bam*HI 切点间的序列，将产生的载体两臂与带有 *Eco*RⅠ 末端的外源片段连接，并通过原生质体转化到酵母细胞中，筛选 Ura^+ 和 Trp^+ 以保证两臂都存在。另外，所使用的宿主带有赭色突变，当载体带有插入顺序时，菌落成红色。

第六节 核酸的序列分析

一、DNA 核苷酸序列的测定

测定 DNA 核苷酸序列的方法现在有两种，一是化学法，又叫 Maxam-Gilbert 法；另一种是双脱氧法，又叫酶法、Sanger 法或链终止法，两种序列分析方法的原理在本书第三章中已经

做过介绍。这两种方法都是要产生标记的不同长度的成套 DNA 片段，各个带有标记片段的起点相同，终点不同，使 DNA 链中的每个核苷酸处都有 DNA 的断裂或终止，通过聚丙烯酰胺电泳，能够将相差一个核苷酸的 DNA 片段分开，用适当的检测手段读出核苷酸序列，再通过计算机分析得到一个完整的 DNA 序列。

（一）双脱氧法分析 DNA 核苷酸序列

Sanger 发明了使用双脱氧核苷酸分析 DNA 序列的方法，其原理是在体外合成 DNA 的同时掺入使 DNA 链合成终止的抑制剂，与四种脱氧核苷酸按一定比例混合，并以之为底物参与 DNA 的体外合成。通常模板链是一条单链 DNA，通过将一段待测 DNA 片段克隆到 M13 噬菌体（第十一章）或其衍生载体上而得到，DNA 的合成由一个通用引物（universal primer）开始，该引物与 M13 的多克隆位点区相邻的区域互补。因此，无论被克隆的 DNA 片段序列如何，都可以使用相同的DNA 合成引物来进行DNA 序列分析。双链DNA 的序列分析基本原理也是相同的，只是在方法上要加以修正，一般双链 DNA 序列分析每组反应所能得到的序列长度不如单链 DNA 序列分析每套反应所能读出的核苷酸序列长。

通常使用的 DNA 链终止抑制剂是 4 种，2′，3′-双脱氧三磷酸核苷酸（ddATP，ddCTP，ddGTP 和 ddTTP），这 4 种双脱氧核苷酸的 3′位没有 DNA 链延长所必需的羟基，所以在 DNA 合成中当掺入一个 ddNTP 时，DNA 链的延长就被终止。DNA 序列分析的体外合成反应一般是在 4 个试管中分别进行的，每个试管中含有一种 ddNTP，反应底物中还要含有一种放射性标记的核苷酸（常为[^{35}S]-ATP 或[^{32}P]-ATP），在 DNA 聚合酶作用下，使 DNA 模板的互补链能够被合成，并随机地掺入双脱氧核苷酸，使 DNA 链的延长终止。因此，每个反应中就产生了一套长短不同的放射性标记的单链 DNA 片段，这些长度各异的 DNA 片段在 5′-末端都是相同的，而 3′端是以一种 ddNMP 结束的。将 4 种反应混合物在变性条件下进行高分辨率的聚丙烯酰胺凝胶电泳（可以分辨长度相差一个核苷酸的单链 DNA），经放射性自显影，就可以读出被测定的DNA片段的核苷酸序列（图14-13）。由于DNA复制时是从5′到3′进行的，所

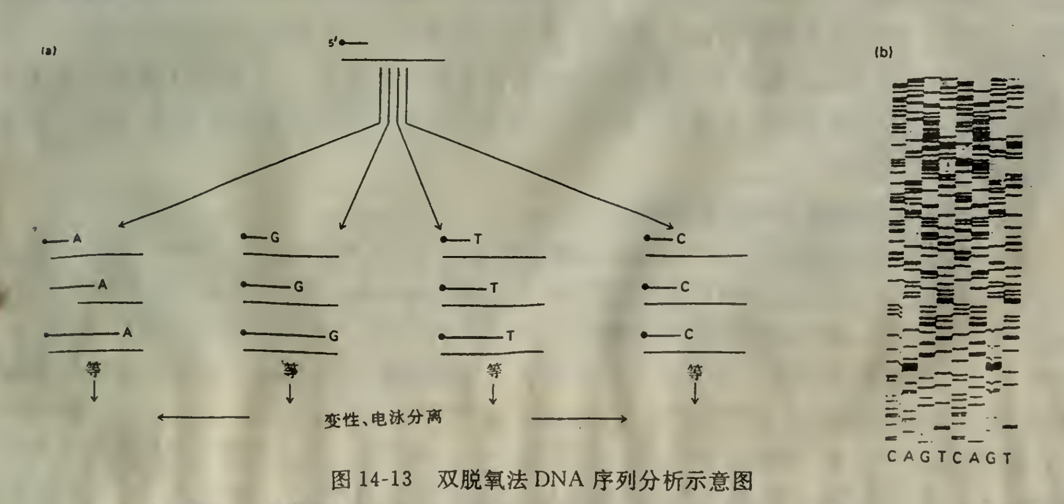


图 14-13　双脱氧法 DNA 序列分析示意图

（a）用一段 5′-末端标记的寡聚脱氧核苷酸引物与待测的单链 DNA 杂交，然后分别进行 4 套反应，在 DNA 聚合酶的作用下 DNA 引物被延长，每个反应中都含有 4 种正常的脱氧核苷酸再加上一种双脱氧三磷酸核苷酸。当 DNA 合成时向 DNA 链中引入一个双脱氧核苷酸时，DNA 链的合成被提前终止。每个反应混合物在高分辨率的聚丙烯酰胺变性电泳胶中分离，经放射性自显影可以读出核苷酸序列；（b）DNA 的核苷酸序列测定的放射性自显影图，反应所用的酶为 sequenase，每个反应做了两次重复，以便于互相证明测序的准确性。

以靠近 5′端的 DNA 链较短，读序列时是从胶的底部读起。

双脱氧法序列分析需要一种酶来催化 DNA 的合成，过去通常使用的是大肠杆菌 DNA 聚合酶 I 的大片段(klenow fragment)，现在多使用经过改造的 T7 DNA 聚合酶(sequenase 和 sequenase 2.0)，经过改造的 T7 DNA 聚合酶可以更均匀地使 dNTP 和 ddNTP 掺入到新合成的 DNA 链中，产生放射性强度较为一致的 DNA 条带。也有人使用耐热的 Taq(*Thermus aquaticus*)聚合酶。这种酶最初用于聚合酶链式反应，而用于核苷酸序列的测定并不十分理想，它所生成的 DNA 带的强度均一性不是很好，但由于它能在较高的温度下催化 DNA 的合成反应，用它对容易形成二级结构的 DNA 片段进行序列分析较为合适。

有些 DNA 在某一片段中含有大量的 GC，容易形成发夹形二级结构，使得这一区域的核苷酸序列难以测出。为解决这一问题，可以使用更加严格的变性电泳条件，例如在电泳胶中加入甲酰胺，或将电泳时凝胶的温度提高，如果通过改变电泳条件不能解决问题时，可以使用 dGTP 的类似物代替 dGTP，常用的类似物有 dITP 和 7-脱氮 dGTP(7-deaza-dGTP)。使用这些类似物所得到 DNA 序列分析图谱质量偏低，因此只有在改变电泳条件无效的情况下才使用 dGTP 的类似物。

在分析 DNA 序列分析的电泳结果时，过去都是将直接从放射性自显影图谱上读取的数据人工地输入电子计算机，当序列较长或数据较多时很容易发生错误，现在有一种自动化数据输入设备(graf bar)，将其放在光源匣的上面，它有一种像笔一样的装置，当用其尖端向每一条带按下时，可以使记录器记下各个电泳道中每一个条带的位置并输入到计算机中，这样即方便了序列分析数据的阅读，又大大减少了出错的机会。

双脱氧法 DNA 序列分析现在已经成了普通实验室都可以掌握的常规手段，与最初相比也有了很多改进，使得一次分析能够比较容易地测出 1 000 bp 以上。标记的方法一般使用随机标记法，即将一种放射性标记的或荧光标记的三磷酸脱氧核苷酸按一定比例掺入底物中，使之能随机地掺入到新合成的 DNA 链中去。这样就增加了灵敏度。由于 DNA 测序多为重复性的操作，人们在不断地探索各种自动化的途径，目前已能将 DNA 序列分析部分地自动化，一些公司推出了体外合成反应和产物分离数据读取的自动化仪器设备，大大提高了测序的速度和准确度，扩大了其应用范围。自动化分析 DNA 序列的仪器和药品目前还很昂贵，因此一般实验室还是以手工测序为主。

(二)DNA 核苷酸序列的化学法测定

用特异的化学裂解法测定 DNA 的核苷酸序列是由 Allan Maxam 和 Walter Gilbert 首先研究成功的。用这一方法进行核苷酸序列测定，首先要制备末端标记(通常是用放射性^{32}P 进行标记)的单链 DNA。可以用多核苷酸激酶将[γ-^{32}P]-ATP 的[γ-^{32}P]磷酸引入双链 DNA 的 5′-末端，再用限制性内切酶从一端切掉一小段并将切下部分除去，就得到了放射性标记的单链 DNA(图 14-14)。用适当的化学试剂处理标记的单链 DNA，使标记的 DNA 在 4 种核苷酸中的一种核苷酸处断开。这一处理的条件是至关重要的，一定要使每个分子断裂的次数平均少于或等于一次。这样就得到了从放射性标记物到 DNA 链中每个特异性核苷酸处的各种长度的放射性 DNA 片段的群体。

假如 DNA 的顺序是 5′^{32}P-ACGTACGT-3′，那么在 4 种核苷酸处分别断裂后就产生了 4 组放射性的群体：

第一个反应，在 A 处断裂^{32}P-ACGT

第二个反应，在 C 处断裂^{32}P-A 和^{32}P-ACGTA

第三个反应，在 G 处断裂^{32}P-AC 和^{32}P-ACGTAC

第四个反应，在 T 处断裂^{32}P-ACG 和^{32}P-ACGTACG

将这 4 组片段进行聚丙烯酰胺凝胶电泳分离，就可以将不同长度的相差为一个核苷酸的 DNA 片段分开。

常用的使 DNA 特异性断裂的化学方法如下：

磷酸二甲酯能破坏嘌呤，它使鸟嘌呤在 N-7 位和腺嘌呤在 N-3 位甲基化，甲基化嘌呤的糖苷键在中性 pH 条件下加热断裂，使嘌呤碱基脱掉，再在碱性条件下加热就使骨架断裂及糖基消除。因为鸟嘌呤甲基化的速度远比腺嘌呤快，所以这样处理主要造成 DNA 在鸟嘌呤处的断裂（图 14-14 中的 G 道）；相反，甲基化腺苷的糖苷键却不如甲基化鸟苷的糖苷键稳定，因此甲基化以后如用稀酸处理主要放出腺嘌呤（图 14-14 中的 A 道）。

胞嘧啶和胸腺嘧啶可以在肼作用下分解，骨架由六氢吡啶断裂。六氢吡啶可以从 DNA 链上去掉肼反应产物，并消除磷酸酯。这些反应使 DNA 在胞嘧啶和胸腺嘧啶处的断裂（C+T 道）。为了区别胞嘧啶和胸腺嘧啶的断裂位点，可以将肼解反应在 2 mol/L NaCl 存在的条件下进行，这时胸腺嘧啶的反应受到抑制，使得 DNA 链断裂主要发生在胞嘧啶处（C 道）。

用 X 射线胶片对电泳胶进行放射性自显影，就可以从胶片上读出所要分析的 DNA 的序列。通过比较凝胶上 G，A，C+T 和 C 道的放射性区带，就可以读出 DNA 的顺序（图 14-14），因为最短的片段迁移速度最大，所以距放射性标记近的一端（5′端）从胶的底部读起。

图 14-14　DNA 序列的化学法测定

（1）首先制备末端标记的单链 DNA 片段，将 4 个相同的含有末端标记的 DNA 片段样品做化学处理；（2）使它们分别在 A，G，C+T 或 C 处断裂，化学反应的程度要控制得使每个标记的 DNA 片段只断裂一次，使被断开的标记的 DNA 在一端带有标记，另一端为化学处理的断点；（3）将 4 个样品分别进行凝胶电泳分离和放射性自显影，可以使各个电泳道中的核苷酸断裂点被读出来；（4）A 道和 G 道中的带可以分别地直接读 A 和 G，同时出现于 C+T 道和 C 道的放射性自显影带读作 C，而只在 C+T 道中出现的带被读做 T。

化学断裂法经常用于分析较短的 DNA 链，例如检测人工合成的寡聚核苷酸和分析蛋白结合区的核苷酸序列。有时一些 DNA 序列由于二级结构的存在很难用双脱氧法测准，这时使用化学法序列分析比较合适。有一种改进的化学测序方法，也可以用于很大的 DNA 片段的核苷酸序列分析[PNAS(1984)81:1991～1995]。

（三）待测 DNA 片段的亚克隆

一般测序的 DNA 片段都比较长，不能通过一次测定反应就能够得到所要的全序列。因此，需要对待测的 DNA 片段进行亚克隆，即将一个完整的 DNA 片段被分成若干个有一定程度重叠的较小的 DNA 片段，并使之能在适当的载体（如 M13）上得到扩增，这一过程称为 DNA 片段的亚克隆过程，每一个较小的 DNA 片段都是 DNA 大片段的一个亚克隆。将测得的每一个亚克隆的 DNA 序列进行计算机分析，根据亚克隆之间的重叠序列排出各个亚克隆的前后次序，最后排定整个 DNA 片段的核苷酸顺序。

用双脱氧法测定DNA序列一般是将待测定的DNA片段克隆到M13mp噬菌体或其衍生载体上，用下面4种手段进行全序列的分析，一是根据测出的DNA核苷酸序列合成DNA片段内部的寡聚核苷酸引物，逐步完成全序列的测定。这一做法的缺点是大的DNA插入片段在载体上可能不太稳定，而且合成新的引物需要花去大量的经费和较多的时间。第二种做法是将DNA克隆在靠近通用引物的一端进行不同程度的缺失处理，然后用通用的M13引物进行序列分析。这一方法的缺点之一也是比较大的亚克隆可能不太稳定，缺点之二是可能需要花很长时间才能得到最佳状态的亚克隆。第三种方法是用限制性内切酶将待测定的DNA切成较小的片段，再分别进行克隆和序列分析，但有时在DNA片段的某些部分难以找到很方便的限制性内切酶位点。第四种方法是将DNA随机降解成较小的DNA片段(常用超声波方法)，再将这些降解的片段进行克隆和测序。这种方法在一开始需要对实验条件进行一些摸索，以后的步骤比较快。对一段DNA进行序列分析，往往要用以上提到的两种以上方法的结合才能完成。

二、RNA 的序列分析

远在DNA序列分析方法出现以前，RNA序列分析就可以用碱基特异性核酸酶进行了，但随着方便快速的DNA序列分析方法的出现，RNA序列分析就显得比较慢且不十分方便。尽管现在测定RNA序列已经采用了一些DNA核苷酸顺序测定的技术和方法，但要直接测定一个RNA分子的核苷酸序列可能需要使用几种不同的方法才能完成。

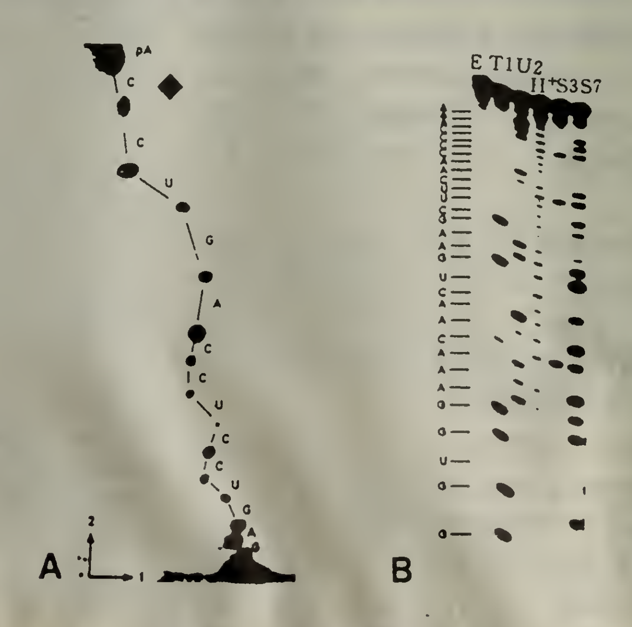


图 14-15　RNA 序列分析

A. 双向迁移率改变法分析15个碱基的RNA片段的核苷酸顺序。用P1核酸酶部分降解以后，用高压电泳(第一向)和同系层析(第二向)，15个核苷酸的排列顺序可以根据已知的迁移速率数据被分析出来，5′-末端标记的寡聚核苷酸是从胶的顶部(5′端)向底部(3′端)读起的。◆为指标剂二甲苯蓝的位置；B. 部分酶解法分析RNA序列。RNA的片段在8 mol/L尿素存在的条件下被碱基特异性的核酸酶部分降解后，用含8 mol/L尿素的20%(W/V)聚丙烯酰胺电泳分离。E：没有外加的核酸酶；T1：T1核酸酶；U2：U2核酸酶；H^+：0.1 mol/L H_2SO_4；S3：pH 3.5时金黄色葡萄球菌核酸酶；S7：pH 7.5和10 mmol/L $CaCl_2$时金黄色葡萄球菌核酸酶。5′-末端标记的寡聚核苷酸在读序时从底部(5′)向顶部(3′)读起。

(一)RNA 的纯化

为避免非特异性的降解，RNA 的纯化要尽可能快地在变性条件下进行，所有的材料和溶液都要严格消毒。对于小分子的 RNA 分子(如 tRNA，5S rRNA，snRNA 等)进行纯化，最有效的方法是双向聚丙烯酰胺凝胶电泳或单向变性电泳与非变性电泳结合。而对分子量大的 RNA 分子如病毒的基因组 RNA 或大的 rRNA，不能直接从细胞中分离，从纯化的病毒粒子或核糖体中进行分离纯化效果要好得多。这些分子量较大的 RNA，也包括 mRNA 通常都经过反转录合成 cDNA，然后分析其 cDNA 核苷酸序列。

(二)RNA 的末端标记

RNA 的序列分析中，对末端的标记是必不可少的。要对 RNA 的 5′端进行放射性磷酸标记，一定要有一个 5′-羟基，帽子结构要用烟草酸性焦磷酸酶和磷酸酶除去，5′端磷酸化的 RNA 要经碱性磷酸酯酶脱磷，并将此酶失活。5′端脱磷的 RNA 可以用高比活性的 T4 多核苷酸激酶(PNK)磷酸化，PNK 可以将[γ-^{32}P]-ATP 的[γ-^{32}P]-磷酸转移到 RNA 的 5′-末端。对 RNA 的 3′-末端进行标记一般用 T4 RNA 连接酶将[5′-^{32}P]-pCp 连接到 RNA 的 3′-末端。

(三)RNA 序列测定

测定由 15 个以下核苷酸组成的序列可以在末端标记后进行部分碱水解或 P1 核酸酶部分降解，再通过双向电泳分离(醋酸纤维素高压电泳和 DEAE 纤维素板同系层析)将不同长度的带有标记的寡聚核苷酸分离开，再根据已知的迁移参数(参看酶学方法 Methods in Enzymology. 59: p58～109.)确定出寡聚核苷酸序列(图 14-15A)。

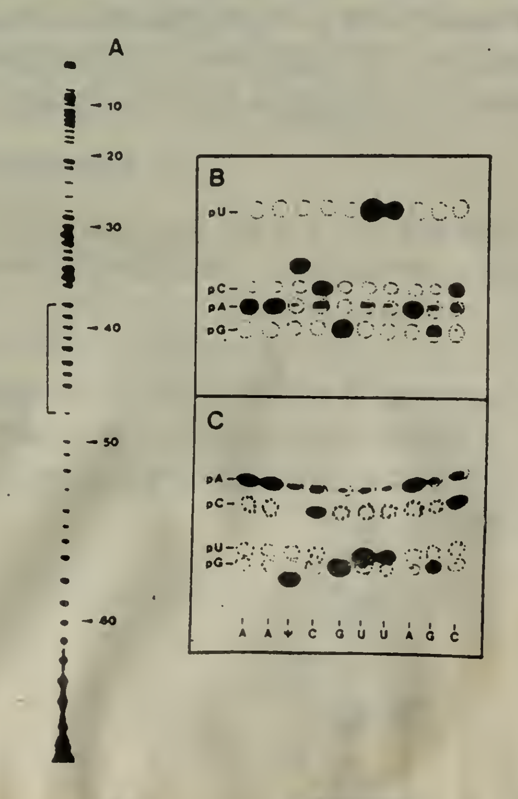


图 14-16　tRNA 的序列分析

A. 将 tRNA 用沸水浴部分降解，降解片段用放射性磷酸进行 5′-末端标记，再用 12.5%聚丙烯酰胺电泳(加 8 mol/L 尿素)分离，将每条放射性条带所对应的 RNA 片段分别洗脱回收，再用 P1 核酸酶彻底降解，用薄层层析鉴定被末端标记的核苷酸 B、C；B、C. 纤维素薄层层析鉴定 5′-末端标记的核苷酸。

末端标记的 RNA 分子也可以用碱基特异性的 RNA 酶的部分降解来进行核苷酸序列的测定。用这些酶处理后，末端标记的 RNA 分子经过高分辨率的聚丙烯酰胺凝胶电泳分离，就可以像 DNA 序列的化学法测定一样，读出 RNA 的核苷酸排列顺序。碱基特异性好的RNA 酶有 RNA 酶 T1(G 专一性)和 RNA 酶 U2(A 专一性)，但目前还没有找到专一性地作用于 C 或 U 的 RNA 酶。PhyM RNA 酶作用于 U 和 A，蜡状芽胞杆菌(*Bacillus cereus*)RNA 酶专一性地作用于 C 和 U，金黄色葡萄球菌(*Saphylococcus aureus*)核酸酶在 pH 7.5 时特异性地作用于 A 和 U，但这一活性和 PhyM RNA 酶不同的是它作用于 U 和 A 前面的磷酸酯键，使得断裂后带有放射性

标记的部分在电泳图谱中向前移动一位(图 14-15 B)。在 pH 3.5,且没有 Ca^{2+} 存在的情况下,金黄色葡萄球菌 RNA 酶表现为嘧啶专一性,有些像蜡状芽胞杆菌核酸酶,但对嘧啶与腺嘌呤之间的键有更强的专一性。部分酶解法分析 RNA 的核苷酸序列具有一定的优点,即速度较快,反应不受 RNA 分子二级结构的影响,并且现在已经有商品化的试剂盒投放市场。其缺点是不能测定出 RNA 分子中的稀有碱基。对于带有修饰基团的 RNA,可以通过对碱基特异性修饰的方法来测定碱基顺序。其原理是将 RNA 的 3′端用放射性基团进行标记,再用碱基特异性试剂将核苷酸的碱基去掉或将碱基的芳香环打开,使 RNA 在受损伤部位的肼解。通过电泳分离后,得到核苷酸的序列。在用酶解法测定 RNA 碱基序列时,一般需要一个碱基阅读梯度作为参考,即随机地在每个碱基处降解,可以通过部分地酸水解或碱水解来得到(图 14-15 B)。

对一些不适合作 5′-末端标记的 RNA(如 tRNA 或 5 S rRNA),可以采用 Stanley 和 Vassilenko 发明的方法来完成序列分析,这种方法是将 RNA 用化学方法随机地降解,可以将 RNA 溶于蒸馏水中,在沸水浴中加热 30 s 左右,使每个分子平均断裂一次(要靠实验得到最佳条件),产生一系列不同长度的 RNA 片段,且又不受 RNA 二级结构的影响。然后用[γ-^{32}P]ATP 和 PNK 将带有 5′-末端羟基的 RNA 片段进行放射性标记,经过电泳分离后,便可纯化得到每一种 5′-末端标记的 RNA 片段,再将这些纯化的片段完全水解,进一步鉴定其 5′-末端的核苷酸,就可以推算出整个 RNA 分子的碱基顺序(图 14-16)。

三、基因转录起始位点的确定

有时确定 mRNA 转录的 DNA 起始位点是很重要的,有两种方法可以完成这一工作:S1 内切酶法或引物延伸技术。首先要根据序列分析数据找到 DNA 分子上含有转录起始位点的区域,再用限制性内切酶将这一片段(约 100 个核苷酸)切掉并分离出来,将 5′-末端进行放射性标记,然后与 mRNA 进行分子杂交,再用 S1 内切酶将未杂交部分的单链切掉,最后通过变性电泳得到标记的单链 DNA 片段,经过序列分析就可以确定起始位点的确切位置(图 14-17 b)。引物延伸技术的原理与此相似,用一段标记的与 mRNA 靠近 5′端互补的寡聚核苷酸,使之与 mRNA 杂交,然后用反转录酶合成 mRNA 的 5′端互补 DNA,变性后电泳分析mRNA的转录起始位点(图 14-17 a)。

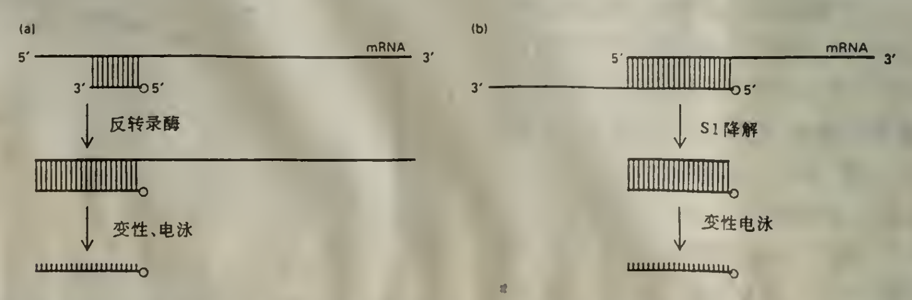


图 14-17 DNA 中 mRNA 所对应的起始位点的测定

(a)引物延伸法:合成一个与 mRNA 完全互补的寡聚核苷酸并进行放射性末端标记,当与 mRNA 杂交之后,用反转录酶合成 mRNA 的 cDNA。变性后得到放射性的单链 DNA,可以测得在 DNA 中对应于 mRNA 第一个核苷酸的位置;(b)S1 内切酶法:制备一段末端标记的单链 DNA(约 100 个核苷酸),其序列是已知的,且对应于 mRNA 5′-末端及其上游区。当这段 DNA 与 mRNA 完全杂交之后,用 S1 内切酶进行酶切处理,去掉没有杂交的单链部分,经变性后,就得到了放射性标记的单链 DNA,可以由此得知 mRNA 起始位点所对应的 DNA 的位置。

第七节　聚合酶链式反应技术及其应用

自1985年Kerry Mullis发明DNA聚合酶链式反应(polymerase chain reaction，PCR)技术以来，这一技术已经在很大程度上代替了传统的DNA克隆方法。利用PCR技术能够从复杂的DNA分子群体中选择性地复制一段特异的序列，使某一DNA片段得到特异性的扩增。同时这一技术也在很多方面得到了广泛的应用。

一、基本原理

如图14-18所示，将提取的总DNA用限制性内切酶切成片段(或用cDNA)，变性后使之成为单链，变性后的DNA将作为DNA聚合酶链式反应的模板。合成分别与待扩增的双链DNA片段两条链的3′端互补的寡聚核苷酸引物(一般每个引物多于15个核苷酸)，然后在50～60℃的温度下将过量的引物加入到少量的DNA样品中，这时DNA保持单链状态，而合成的特异性的引物能够与其互补序列杂交，这些杂交上的寡聚核苷酸将作为引物，参与合成模板DNA的互补链。加入三磷酸核苷酸(dNTP)和耐热的DNA Taq聚合酶后，DNA合成反应开

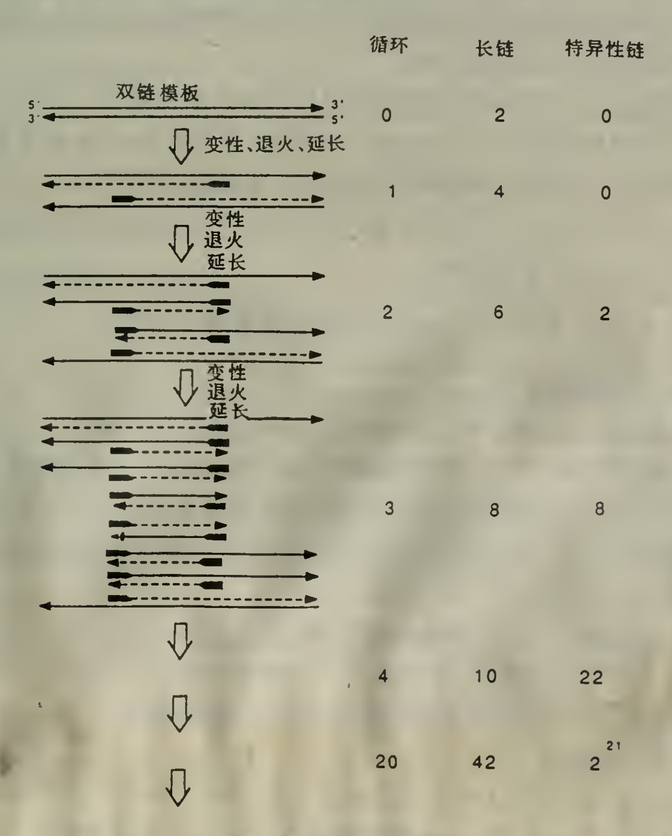


图14-18　聚合酶链式反应示意图

双链DNA分子(细长箭头)经过热变性后，用两个与待扩增DNA片段相邻的核苷酸序列互补的寡聚核苷酸引物与变性DNA一起退火，再用一种DNA聚合酶使引物延长，变性-退火-延长构成了一个DNA扩增反应循环，像这样的循环可以重复多达40次，使特异性DNA片段得到扩增。

始，Taq 聚合酶能在高达 72℃的条件下进行反应。当反应完成时，将反应混合物加热到 95℃使新合成的 DNA 双链变性，当温度下降以后，由于过量引物的存在，另一轮的反应又可以进行。这种合成-变性-退火-合成的循环可以重复多次，使所需要的 DNA 片段得到特异性的扩增，在理想的条件下，每一次循环使两个引物位点之间的 DNA 序列增加一倍。

二、PCR 技术的问题及改进

第一个问题当 DNA 模板的量太少（如小于 10 个拷贝）时，有时反应失败，产生非特异性的扩增，其原因可能是引物和模板之间非特异性的结合或引物之间相互作用引起的。最近有人提出改进的建议，一是高温 PCR，由于多数非特异性的引物延长反应可能是在第一轮反应的加温到变性过程中温度的最低点发生的，所以可以在第一次高温过后加DNA 聚合酶，以提高反应的特异性；或者在第一循环中不加入三磷酸核苷酸，也能起到相同的作用。另一解决的办法可以称为后援 PCR，在模板的量较小时，可以先用一小部分引物进行反应，经过约 15 个循环之后，再加入其余的寡聚核苷酸引物。

这一技术的另一问题是使用 Taq 聚合酶时，由于这个酶没有外切酶活性，在合成中容易发生错误。最近有报道表明，可以通过控制反应的 pH 值和底物的浓度将核苷酸的错误率降低 5 倍（到 $1/10^5$），当要对扩增的 DNA 片段进行基因克隆或进行功能的研究时，这是很有用的，只是不能充分利用 Taq 酶的反应活性。

PCR 技术应用的另一限制因素是被扩增片段的长度一般在 2.0 kb 以下时效果才最好，当要扩增的 DNA 片段长度较大时，会出现一些问题，可能扩增产物的背影较强，或扩增出的产物 DNA 片段的电泳带不整齐，或没有扩增发生。最近有报道（Nucl. Acids Res. 20：623，1992）说，使用 pH8.4 的不含 KCl 的 Tricine 缓冲液，作者可以重复性很好地用 Taq DNA 聚合酶来扩增得到 6.3 kb 的 DNA 扩增产物。

三、PCR 的应用

（一）引入新的核苷酸序列

新的核苷酸顺序可以加到 PCR 引物的 5′-末端，或者通过改变 PCR 引物的核苷酸序列而获得。新的核苷酸顺序可以是一个新的酶切位点，这将大大方便克隆特异 DNA 片段的操作；也可以是特异的突变，突变的引入对于结构和功能的关系的研究是一种很好的手段；也可以通过新的核苷酸顺序的引入，将基因表达的调控序列如启动子，或翻译的起始位点等加到基因的上游，从而使 PCR 产物在体外系统中得到表达。

把碱基加到引物的 5′端是最简单的方法，新顺序的加入对 PCR 无任何影响，目前长达 45 bp 的 5′端加入顺序已有报道。若要在 PCR 引物中引入突变，也应尽可能将突变位点设在引物的 5′-末端，这样才能最大限度地降低突变点对 PCR 的影响。

（二）重复序列的扩增

PCR 技术已被用来扩增 DNA 中的重复片段，所使用的引物核苷酸有与重复序列相邻的序列，也有在重复片段内的序列。典型的例子是人的 Alu 序列家族的 PCR 扩增，由于 Alu 序列分布在人类基因组中的不同区域，而且是人类基因组所特有的顺序，因此，Alu 序列的 PCR 可用于从非人类 DNA 背景中鉴定、扩增和分离人类 DNA 序列。

构建指纹图谱是重复 DNA 片段 PCR 的另一用途。以重复序列作为引物进行的 PCR，不

同的个体之间由于基因组结构具有足够的天然变异，而导致扩增重复序列之间的不同DNA片段对其效率不同，最终结果是产生了可区分的不同个体特有的指纹图谱。

（三）锚定 PCR

锚定PCR的目的在于分离与一段已知序列相邻的未知序列，潜在的应用前景有鉴定复杂基因组的内含子外显子结合部、重组过程中整合位点、引起人类疾病的DNA断点等。还可以分离与固定序列相邻的DNA片段，简化克隆和染色体行走步骤等。人们已经设计和成功地使用了一些方法，其中包括从已知序列到人工加上去的寡聚核苷酸的PCR，从已知序列到天然的或人工连接的同种核苷酸序列的PCR，环化DNA片段的PCR和Alu序列的PCR等。还可以通过一些辅助措施进行这些工作，如在寡聚核苷酸的5′端用生物素进行标记，再用亲合素包被的磁珠捕捉扩增的DNA片段；或在5′端加入一段双链DNA的蛋白质结合序列，用固定化的特异性DNA结合蛋白来纯化扩增后的双链DNA。

还有一种单位点PCR，从一个已知位点的寡聚核苷酸向一个人工连接上的寡聚核苷酸引物进行PCR扩增（图14-19），连接上的引物又是PCR所用的一个引物，只有已知序列的引物被延长并合成到人工连接的引物位置后，另一个引物才能被延长，从而达到特异性扩增的目的。

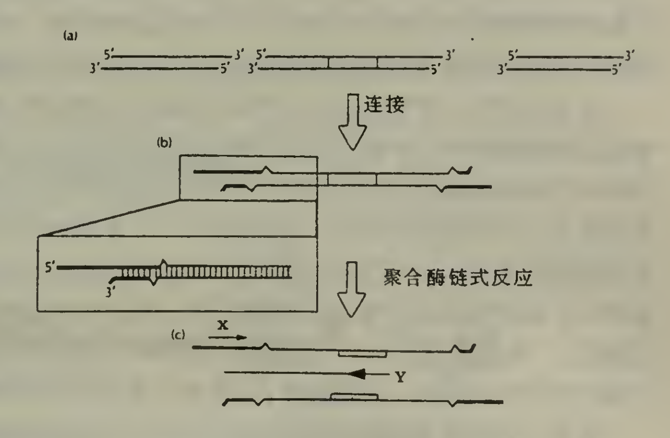


图14-19　与已知序列相邻的未知序列的单位点扩增

（a）用限制性内切酶将DNA降解，此内切酶在已知序列内无切点；（b）在连接酶的作用下将一对核苷酸引物加在新切出的DNA片段的末端，这一对连接上去的寡聚核苷酸并不完全配对，即较短的一条在其3′端不能与较长的一条形成互补氢键；（c）以连接上去的寡聚核苷酸和相邻的已知序列内的一个引物进行DNA聚合酶链式反应，由于连接上去的一段寡聚核苷酸引物不完全配对，使得X方向进行的DNA合成，只有在Y方向的DNA合成完成以后才能够进行，保证了PCR引物的特异性。

（四）再生 PCR

在分子生物学研究中有时受到所能得到的DNA量的限制而使得一些很有用的技术方法不能使用，PCR扩增可以解决这一问题，将整个DNA群体中各个片段的两端都连接上寡聚核苷酸，再用同样的寡聚核苷酸作为引物进行PCR，使片段混合物得到扩增，这一过程为再生PCR（regenerative，PCR）。这一方法已成功地应用于人的cDNA差减筛选等工作。需要注意的问题是长度不同的DNA片段可能不成比例地被扩增。

（五）定量 PCR

一般通过PCR进行DNA扩增的效率有很好的重复性，因此这一方法可以用来估计模板DNA的浓度，但要有一个已知含量的对照模板，与未知模板一起进行扩增。

（六）突变的检测

应用较广泛的是检测常见的遗传病，用PCR合成引起病变处的DNA序列，再用病变的寡聚核苷酸进行点杂交，单个的核苷酸变化能够被检测出来。对于那些多处突变引起的病变可以合成多个与引起病变序列互补的核苷酸，分别点在杂交的膜上，再用标记的PCR产物与膜进行杂交，根据杂交反应的阴性或阳性来确定是哪处碱基变化引起的遗传病，这种方法称为反

向点杂交。最近在这方面有两种方法上的改进，一是根据已知遗传病的核苷酸顺序的改变合成正常的和突变的PCR引物，在PCR反应的底物中加入荧光标记的核苷酸，根据反应产物中有否荧光出现，确定在扩增所用的引物中有否碱基变化；另一个进展是利用完全配对的PCR引物，来检测PCR产物中可能的核苷酸序列改变，反应完成后将变性的放射性标记反应产物用非变性的聚丙烯酰胺电泳进行分析，单个的核苷酸变化会使电泳图谱发生变化。

（七）直接分析PCR产物的核苷酸序列

要测定PCR产物的核苷酸序列，一般不需要将PCR产物克隆到一个序列分析载体上，可以直接用PCR的引物进行序列分析。目前有几种方法可以用来测定PCR产物的核苷酸序列，其一是用^{32}P标记的引物，而不是将放射性同位素随机地掺入新合成的DNA链中去，这样可以增加放射性信号强度，降低背景条带的强度；但这种做法很难测出大于200 bp的序列，主要是由于互补的两条模板单链很容易形成双链，而序列分析的引物被排斥在外。解决这一问题有两种办法，一是向序列分析反应中加入10%（V/V）的二甲亚砜，以阻止互补双链的复性；另一方法是在进行序列分析反应时，使用多次热循环，即在ddNTP存在的条件下进行多个循环的PCR扩增。

由于在DNA序列分析中，DNA的第二条链常常对测序反应有干扰，所以人们想出了一些办法来将两条DNA链在进行序列分析反应以前分离开来，其一是将DNA双链在变性电泳中分开，再将两条链分别回收，这一方法对碱基含量较偏的DNA效果最好。一个最新出现的方法是在一个PCR引物的5′-末端引入一个生物素基团，再用带有亲合素琼脂糖柱纯化带有生物素的单链DNA，洗脱后就得到了单链的DNA产物，便于DNA的序列分析。第三种将两条DNA链分离的方法是用外切酶Ⅲ将一条DNA链降解，在PCR扩增以前，先将一个PCR引物的5′-末端磷酸化，由于外切酶Ⅲ只降解5′端带有磷酸基团的DNA链，就使两条互补的扩增产物DNA链的一条被降解，而另外一条链可以用来作为DNA序列分析的模板。

还有一种不对称扩增法可以用来制备DNA序列分析的模板，这种方法在直接分析PCR产物核苷酸序列中用的是最多的。其原理是在DNA扩增时加入的两种PCR引物的量是不同的，其中的一种引物在PCR扩增前10～15个循环中就被用完了，而在以后的多次循环中，只有另外一种引物被延长，PCR的产物的大部分为一条DNA链，在以后的测序反应中所受到另一条链的干扰就小得多了。

（八）微生物PCR

很多人正在利用PCR技术试图发现某种微生物的存在。有一种做法设计得比较巧妙，将免疫反应与PCR扩增结合起来，先将抗甲型肝炎病毒的抗体固定在微型滴定板上，然后与含有甲肝病毒颗粒的样品接触，经过洗涤后，得到甲肝病毒，再用特异性的引物对所得到的病毒的DNA进行PCR扩增，就证明了甲肝病毒的存在。

也有人将PCR技术用于细菌的鉴定和分类上，细菌中与rRNA基因相邻的DNA序列因种而异，可以对这些因种而异的核苷酸序列进行PCR扩增和序列分析，达到对不同细菌进行鉴定的目的。

（九）合成PCR

有人对人工合成的234碱基的序列进行了PCR扩增，先合成一条完整的单链DNA，再合成两端的寡聚核苷酸引物，最后用PCR技术进行扩增。结果证明没有化学反应和酶反应引入的错误。这种方法要比化学法合成两条互补单链或先合成两条相互部分重叠的单链再用酶法补齐的办法快而且便宜，尤其是在DNA产物用量较大的情况下更为有用。

第八节　基因的表达和改造

DNA 的同源重组是一个生物界的普遍现象，这一研究领域除了本身所具有的理论意义之外，其应用前景对人类更具有莫大的诱惑力。分子生物学技术研究的直接应用前景就是应用于生产实践和临床医疗，通过一定的技术手段将某些基因加以改造，有时需要对基因的产物的结构进行一定修饰，在有些情况下需要增加某基因的活性，相反，在另外一些情况下，却要降低某些基因的活性，甚至有时需要将某些基因完全失活，才能满足人的愿望。但基因改造是一个复杂的过程，目前还存在着一些难以克服的障碍，例如，有些生物难以用外源 DNA 进行转化，转化后外源 DNA 在染色体上随机地与内源 DNA 重组，造成基因组功能上的紊乱等等。为此，人们也正在逐步探索有关的技术原理和完善基因改造的方法。

对原核细胞进行基因改造比较简单，例如大肠杆菌，可以筛选出目的基因的突变型，再将改造过的基因引入到突变型中；而真核细胞的基因改造就要复杂得多，下面介绍几种真核基因改造中常用的方法。

一、体外定点突变

（一）基因调控单元的突变

当一个基因被克隆并能引入到它原来的细胞中以后，就可以通过一系列的突变手段找出基因的调控单元，开始鉴定调控单元靠选择突变型，现在可以通过一些方便快捷的手段来进行。要确定一段调控单元，需要对基因的调控区进行缺失、插入、预定位点的碱基对置换等突变手段。最容易做的是两个内切酶位点之间的 DNA 片段缺失。对一个质粒进行部分酶切，可以分离得到大量的只缺失一个内切酶片段的酶切产物，进一步将这些纯化的 DNA 连接成环状，其中相当一部分还会有生物活性。要制造比较短的缺失，可以将质粒 DNA 在一个内切酶位点上切开，使之成为线状的，然后使用核酸外切酶对质粒 DNA 两端进行缺失处理，如图 14-20。(a)DNA 缺失：可以将一个 DNA 克隆中的两个内切酶位点之间的 DNA 片段全部去掉，也可以从一个内切酶位点切开以后，用外切酶 *Bal*31 进行缺失处理，*Bal*31 可以从线性化的 DNA 两端逐步地切掉一到数个碱基对。所需要的 DNA 缺失克隆可以从一系列的不同程度的缺失克隆中筛选出来；(b)在要制造突变的位置设计一个化学合成的寡聚核苷酸引物，使得

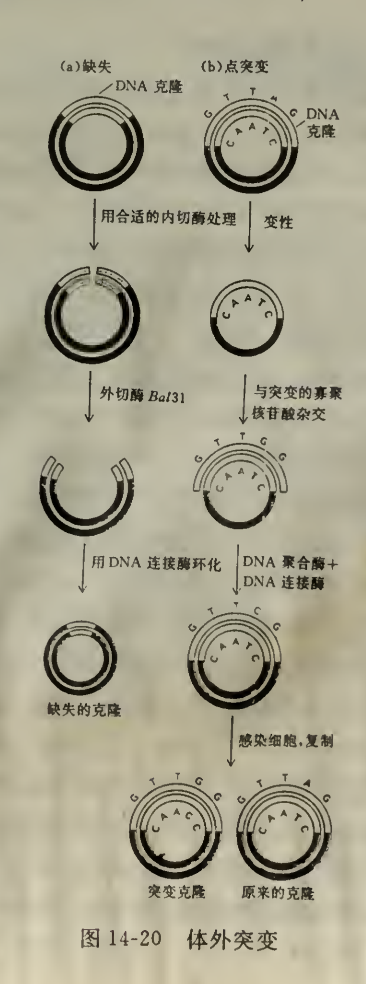


图 14-20　体外突变

这一外来引物与变性后的 DNA 克隆杂交，再用 DNA 聚合酶将引物延长，转化细胞以后，就产生了正常克隆和带有人工引入的突变克隆的混合物，经过进一步筛选就可以得到所需要的突变克隆。

（二）碱基特异性突变

单个碱基对的改变可以使用体外定点突变，以代替过去使用的化学方法。用一段 15～20 个碱基的寡聚核苷酸链，可以在一条 DNA 分子的任何部位进行碱基特异性的改变。这种长度的寡聚核苷酸可以和互补 DNA 链形成非常稳定的 DNA-DNA 双螺旋结构，当寡聚核苷酸含有一到两个非互补碱基时，这种互补的双螺旋结构就不十分牢固，但仍然具有相当的稳定性。在 DNA 聚合酶和 4 种三磷酸脱氧核苷酸存在的条件下，这种不完全互补的寡聚核苷酸可以作为 DNA 链延长的引物，所生长的双链 DNA 除在引物部分以外都是互补的。将这样的环状 DNA 分子引入大肠杆菌后，就会产生两种 DNA，一种是野生型的，另一种是在原来的寡聚核苷酸引物处碱基改变的 DNA（图 14-20）。两种 DNA 分子可以由它们对突变部位的不同寡聚核苷酸的结合紧密程度通过 DNA 杂交来区分。野生型的 DNA 对野生型寡聚核苷酸比对突变型的寡聚核苷酸的结合能力强，而突变型的 DNA 却相反。

二、体内基因的改造

（一）野生型的恢复

大肠杆菌中的真核基因克隆所能提供的信息是有限的，要研究基因的调控和功能，还要将基因引入到真核细胞中去。酵母细胞的转化常通过钙离子（Ca^{2+}）和聚乙二醇处理酵母的原生质体，使外源 DNA 被引入到酵母细胞中，第一个成功的例子是将克隆在大肠杆菌质粒上的酵母 *LEU*2$^+$转化到 *LEU*2$^-$的单倍体细胞中，所有的能够在不含亮氨酸的培养基上生长的细胞都是由于 *LEU*2$^+$被整合到了酵母的含有 *LEU*2$^-$的染色体上，这是通过重组酶类的作用，经过同源交换来实现的（图 14-21）。*LEU*2$^+$基因插入时最容易出现的情况是一次发生在质粒和染色体之间同源序列之间的对等交换，引起整合转化即整个外来质粒所携带的 *LEU* 2$^+$插入到同源染色体中，使得染色体上含有两套基因，一是原来细胞中的 *LEU*2$^-$，还有一套是新引入的野生型，这样转化了的细胞就具有了自身合成亮氨酸的能力，可以在不含亮氨酸的培养基上将其筛选出来。如果发生同源区的双交换，
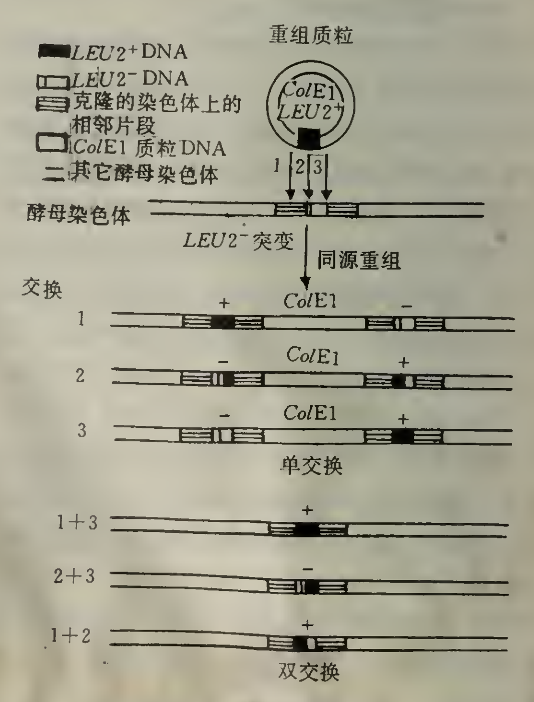


图 14-21　酵母的整合转化

酵母的重组质粒上带有一个野生型的基因（*LEU*2$^+$），酵母染色体上带有一个突变的基因（*LEU*2$^-$），3 个并列向下的箭头表示可能发生同源重组的位置。

就导致外来基因的片段替代原有序列，当序列的替换发生在引起 *LEU*2 基因突变的区域时，*LEU*2$^-$就变成了 *LEU*2$^+$，细胞也具有了合成亮氨酸的能力。

随后人们又成功地用这一方法转化了酵母的一些其它基因，但转化效率很低。这一问题通过使用含有酵母 DNA 复制起始位点的质粒和将引入的 DNA 分子线性化得到了克服。

（二）酵母染色体基因的替代

不仅可以用野生型基因取代突变的基因，整合性转化也可以用来以突变基因代替正常基因。这一基因操作手段首先应用于酵母整合型质粒（YIp）的转化，YIp 中含有缺失、无活性的 *HIS*3⁻基因和一个有活性的 *URA*3 基因，并带有这些基因的相邻序列区。用这样的质粒转化酵母使得半数质粒插入到 *HIS*3 位点，另一半质粒插入到位于另一个染色体的 *URA*3 位点上。在 *HIS*3 位点的插入使酵母染色体带上两个相邻近的 *HIS* 位点，一个野生型的和一个突变无活性的（图 14-22）。这些细胞的进一步生长会导致不时的在两个 *HIS*3 位点之间发生交换，交换的结果使得两个 *HIS*3 位点之间的 *URA*3 序列从染色体上被剪切下来，使细胞重新变为 *URA*3 的突变型。这些具有 *URA*3⁻表现型的恢复子中有些还是 *HIS*3 的野生型，而有些已经变成了 *HIS*3⁻的突变型（图 14-22）。

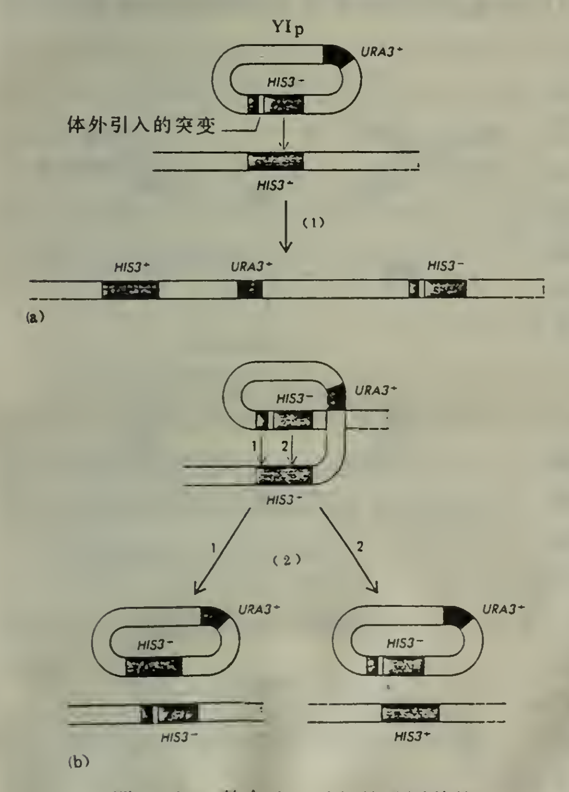


图 14-22　整合重组引起的基因替换

带有一个突变的 *HIS*3 基因和一个野生型的 *URA* 基因可以通过在染色体的 *HIS*3 或 *URA* 位点分离而使酵母被转化。(a)转化发生在 *HIS* 位点附近，根据质粒是否从染色体上被切除，产生转化了的或野生型的染色体；(b)在染色体 *HIS*3⁺位点附近发生的整合，使得两个 *HIS*3 基因在染色体上串行排列，两个 *HIS*3 拷贝之间发生的重组使整合过程逆转，即质粒被剪切下来（染色体丢掉 *URA* 选择标记），这时有两种可能的情况发生，一种是产生带有突变型 *HIS*3 的染色体，另一种是染色体仍然带有 *HIS*3⁺基因的野生型。

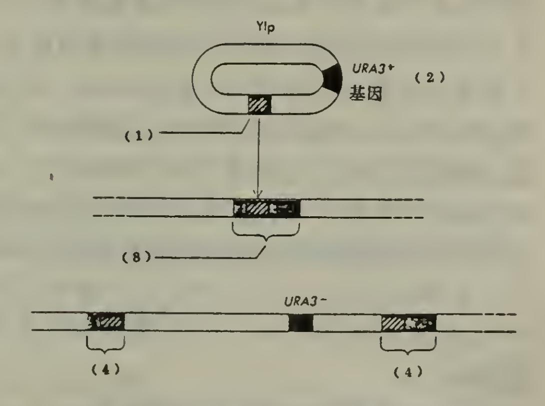


图 14-23　利用整合重组分析染色体基因（肌动蛋白基因）的必要性

(1)酵母肌动蛋白基因的两端被缺失以后，带在酵母的整合质粒（YIp）上；(2)YIp 还带有一个 *URA*⁺基因作为选择标记；(3)经过同源重组后，产生两种突变的无功能的肌动蛋白基因；(4)酵母在没有正常肌动蛋白的情况下不能存活，因此这里所介绍的插入失活实验是在双倍体菌株中进行的（有正常的肌动蛋白存在），半数的减数分裂孢子（即带有失活的肌动蛋白基因的孢子）不能存活，因此肌动蛋白是酵母所必需的。

任何定位于染色体的基因只要有了野生型克隆之后，都可以用更简单的方法将其失活。其原理是在体外将基因的前后两端进行缺失处理，用基因的中间部分转化细胞，当在

细胞中发生交换时，所产生的两种基因拷贝都是有缺陷的，其中一个拷贝没有基因的开始端，另一个拷贝缺少基因的末尾。这一操作首先证明了酵母肌动蛋白的必要性。在这一实验中二倍体的 $URA3^-$ 细胞被转化成 $URA3^+$，用来转化用的质粒上含有一个 $URA3^+$ 基因和一段肌动蛋白基因的中间部分的片段，当在肌动蛋白基因片段处发生交换的 50% 的 $URA3^+$ 细胞（相对于在 *URA3* 基因中发生交换的另外 50%）形成孢子时，只能形成 $URA3^-$ 孢子，因为在这样的 $URA3^+$ 单倍体细胞中肌动蛋白基因是没有活性的（图 14-23）。

（三）酵母基因的获得

酵母的特定基因可以用恢复载体（retriever vector）通过同源交换从正常染色体上获得。用途较广的恢复载体上含有能够在大肠杆菌和酵母中复制的起始位点和能够在两种细胞生长过程中进行选择的标记，同时还要含有一段与目的基因同源的序列。这样的载体构建完成以后，用限制性内切酶去掉中间的同源基因片段，留下基因两端的酵母 DNA 序列，将这一线性的缺失载体引入到含有目的基因的细胞中，当进入细胞核之后，具有高度同源性的两个 3′-末端会从酵母染色体的同源区开始进行复制，结果使细胞中的目的基因被复制到引入的恢复载体上来（图 14-24）。下一步就可以提取质粒 DNA，转化大肠杆菌，进行扩增，得到大量的目的基因。

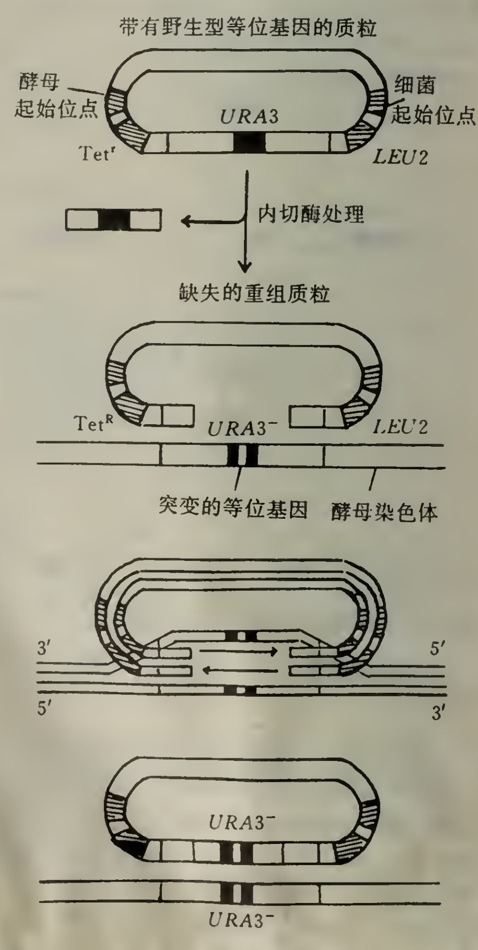


图 14-24 恢复载体的使用

用限制性内切酶去掉恢复载体质粒上的目的基因，这一缺失的载体被引入到带有突变型目的基因的细胞中去，恢复载体进入细胞以后，质粒的末端序列就侵入到与基因相邻的片段，并开始复制与丢掉的基因同源的染色体上的基因序列。

（四）高等真核生物中的同源重组

哺乳动物基因转化可以通过一些常规的基因转化手段来实现，如磷酸钙/DNA 共沉淀、原生质体融合、电穿孔、直接注射等。一些高等植物的基因转化可以通过农杆菌的 Ti 质粒来实现，也可以使用电穿孔或基因枪来进行转化。

哺乳动物中最初进行的基因改造工作都是先向细胞的染色体上引入一个突变的基因，再用另外一个突变的同一基因转化细胞，使外来的基因和已经定位在染色体上的基因发生同源重组，形成一个野生型的基因。引入外源的 DNA 一般使用在与准备改造的基因同源处线性化的，这样可以大大提高基因改造效率。Smithies 等人第一次对天然基因位点（β-球蛋白）进行改造，同源对非同源重组的比约为 1 000 : 1，经过改造的 β-球蛋白能够稳定地进行表达。

植物基因转化可以借助于农杆菌天然的转基因性质，也可以直接对植物的适当组织和细

胞进行转化。一个直接转化成功的报道是对烟草原生质体的两次转化将两个突变的新霉素磷酸转移酶在原生质体中同源重组，产生野生型的卡那霉素抗性基因。首先通过基因转化构建受体细胞系，使细胞系中含有一端缺失的基因，再用另一端缺失的基因转化构建的受体细胞系，作者筛选到了同源重组的野生型新霉素磷酸转移酶原生质体，并且培养出再生植株，野生型性状能够稳定地遗传给后代。

农杆菌进行基因转化的核心部分是质粒编码的 T-DNA（转化的机制尚不清楚），因此人们试图用 T-DNA 作为一个过渡分子，使染色体外的遗传物质能够在细胞核中进行同源重组而达到基因改造的目的。1990 年有利用农杆菌 T-DNA 进行基因改造成功的报道。作者将叶肉细胞原生质体与两个分别带有缺失选择标记的 T-DNA 的农杆菌共培养，两个 T-DNA 在选择标记基因的编码区有 0.55 kb 的同源区，一个在编码区的 5′端缺失，一个在编码区的 3′端缺失，同源重组的原生质体被特异性地选择出来。他们发现约有 1%～4%的转化细胞发生了分子间的同源重组，对 11 个不同的再生株系进行 Southern 印迹和 PCR 分析，证明了有野生型标记基因的存在。有一种可能性是重组发生在农杆菌细胞中，而不是在植物组织中，首先通过质粒结合使同一细菌细胞中含有两种类型的 T-DNA，两个不同的缺陷型的标记基因在细胞内发生重组，产生野生型重组，再将野生型的重组基因转入受体植物组织。这种可能性已被各种对比实验所排除。这一实验证明了 T-DNA 可以用作植物细胞核中同源重组的底物。

三、反义 RNA 技术及其应用前景

继在细菌中发现反义 RNA 对基因表达的调控作用（见第十一章）以后，人们试图利用反义 RNA 来特异性地抑制某些基因的表达。美国科学家用单式疱疹病毒的胸苷激酶（TK）基因蛋白质编码区反向连接在启动子的下游，使之能够转录出反义 RNA（即与 mRNA 序列互补的 RNA），当将带有有义基因与反义基因的质粒注射到 TK^-细胞中时，反义 RNA 抑制了细胞中 TK 酶的表达。后来他们又将 *TK* 基因中 52 个碱基的非编码区以相反的方向连接在启动子的下游，并使之在瞬时和长期地转录反义 RNA，发现 TK 的这一反义 RNA 对所抑制的序列是高度专一的，它只能抑制单疱疹病毒的 TK，而对与之没有对应的有义序列的鸡 TK 没有抑制作用。

1991 年有一项在苜蓿上用反义 RNA 进行应用研究所注册的专利，研究的目的是想通过抑制根组织特异性表达的谷氨酰胺合成酶亚基，使根中谷氨酰胺的水平下降，以缓解对固氮酶活性的抑制作用。作者们将谷氨酰胺合成酶同工酶以颠倒的方向连接在根特异性表达启动子的下游，连接在一个大肠杆菌-农杆菌双起始位点质粒上，质粒上带有 25 bp 染色体上的相邻序列（整合区），以氨甲蝶呤（methotrexate）抗性为选择标记。实验结果证明，反义 RNA 使得根中谷氨酰胺合成酶的活性降低 80%，苜蓿干重增加 50%。

四、用基因敲除技术研究基因表达及功能

基因敲除技术（gene knockout），又称基因靶向技术（gene targeting）是 80 年代末发展起来的，通过人为干预达到抑制或灭活目的基因表达的分子生物学技术。是利用同源重组技术建立基因定点灭活细胞系或基因定点灭活动物。近年来，在生物学诸多研究领域中得到应用，已获得了数百种基因定点缺陷小鼠。这些研究成果使人们对被敲除基因在生物总体水平的功能有了更深入地了解。

该技术的原理是首先用基因重组技术在目的基因片段中插入一外源基因作为筛选标志基因并使目的基因失活。将此“灭活”基因转入小鼠胚胎干细胞(ES 细胞)中，通过细胞内的基因同源重组使灭活的基因取代染色体上原有的目的基因。利用标志基因产物的特殊药物抗性进行筛选和基因型鉴定，可以得到基因定点灭活的 ES 细胞。将重组 ES 细胞注射到小鼠胚胎中，再将胚胎植入假孕母体的子宫内，使其发育成目的基因缺陷的杂合的嵌合体小鼠，最后从子代小鼠中得到纯合子的基因敲除小鼠。通过分析基因敲除小鼠体内单基因缺陷所产生的表型异常来研究基因调控和功能，建立疾病的动物模型以及进行基因治疗等。

第九节　RFLP 和 RAPD 及其应用

RFLP 是限制性片段长度多态性(restriction fragment length polymorphism)的英文缩写。RFLP 是由于不同个体的等位基因之间碱基代换、重排、插入、缺失等引起的。限制性内切酶能把很大的 DNA 分子降解成许多长短不等的较小片段，所产生的 DNA 片段的数目和各个片段的长度又反映了限制性内切酶的切点在 DNA 分子上的分布。对于每一个 DNA/限制性内切酶组合来说，所产生的片段是特异性的，它能作为某一 DNA(或含这种 DNA 的生物)所特有的“指纹”。

当一个基因组比较小时，例如植物叶绿体基因组，只有十几万碱基对，就可以将提纯的 DNA 用限制性内切酶消化成各种长度的 DNA 片段，进行琼脂糖凝胶电泳分离并用溴化乙锭染色以后，可以在紫外光下直接观察多态性(图 14-25)。但对于大分子染色体 DNA，就不能直接观察限制性片段长度的多态性，因为各种长度的 DNA 片段在电泳胶上相互交盖，连成一片，不能进行分辨，要检测限制性片段长度的多态性就要借助于分子杂交手段，用某一标记的 DNA 片段作为探针，与被转到硝酸纤维素膜或尼龙膜上的电泳后的 DNA 片段进行杂交，这样只有与探针有高度同源性的片段能够被检测出来，因此提到限制性片段长度的多态性一般都要通过 DNA 分子杂交(Southern blotting)来检测。

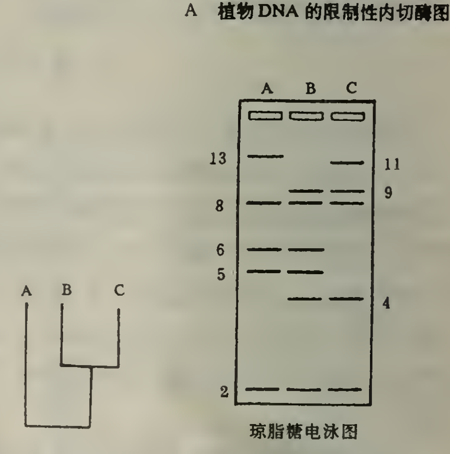


图 14-25　系统发育的 RFLP 分析

右侧上方表示 A 植物叶绿体基因组小部分 DNA 限制性内切酶图谱，右侧下部为植物 A 和植物 B 及植物 C 的 RFLP 形式。如果植物 A 与祖先植物种相比变化很小，那么假设它发生了两个突变，一是在形成植物 B 的过程中，13 kb 片段上出现一个新的限制性切点，产生一个 9 kb 和一个 4 kb 的片段；另一个突变是在形成植物 C 的过程中，一个限制性内切酶位点消失，产生 11 kb 的内切酶片段，而 5 kb 和 6 kb 两条带消失。可以根据这些变化列出系统发育图如本图左侧所示。

10 多年来 RFLP 作为共显性的遗传标记，被广泛地用于很多种生物的分析中，这主要是由于 RFLP 的普遍性以及检测方法的简便性。限制性片段长度多态性在植物遗传图谱的构建上尤其有用。过去跟踪染色体片段和观察它们在配子形成时重组的唯一方法是观察发生在该片段上的基因作用所引起的表现型的变化。通过观察表现型变化，如花色、株高、抗性等，人们可以推测在减数分裂和重组期间各个染色体片段的行为。用这种间接的方

法跟踪染色体片段已构建了几个植物种相当详细的遗传图，但是这一方法非常繁琐，而且要花很长的时间。而 RFLP 提供了一个直接跟踪重组期间染色体片段的方法，而且大大地简化了遗传图的构建过程。人们只需直接看片段本身的 RFLP 标记，而不必看染色体片段上基因存在所引起的表现型变化。因此，可以直接看到植物的基因型，而不是间接地通过基因产生的表现型来推测。

由于在一个作图群体中可以分析大量的 RFLP 标记，所以能够构建一个分离群体的染色体片段图。这些图能表示群体中每一植株的整个基因组中来自父、母本双方的染色体片段的分布，这在遗传育种中是非常有用的，例如在一个回交计划中，目标是导入一个单基因，那么我们要选择的植株应该是：含轮回亲本染色体片段数目最大、同时要保留目的基因的片段。

Alec Jeffreys 等从人的染色体 DNA 中分离出一种中度重复序列作为探针，对人的染色体 DNA 进行 RFLP 分析（称为 DNA 指纹）。他们将总 DNA 用限制性内切酶完全酶解以后，用放射性的探针进行 Southern 印迹（见本章第三节），每个人的 DNA 所产生的阳性杂交条带都是特异性的，而且子女的阳性杂交条带中，有一半与父亲的 DNA 阳性杂交带相同，另一半是与来自母亲的阳性杂交条带相同（图 14-26）。用这种中度重复序列作为探针，对不同的人的 DNA 进行指纹分析，产生完全相同的图谱的机会小于 6×10^{-9}，因此这一方法可用于法医鉴定，血缘关系的识别等。这一实验方法所提供的证据已于 1988 年得到英国法庭的认可。由于英国有着严重的移民问题，DNA 指纹技术的应用，为新移民血缘关系的鉴别提供了令人信服的技术手段。

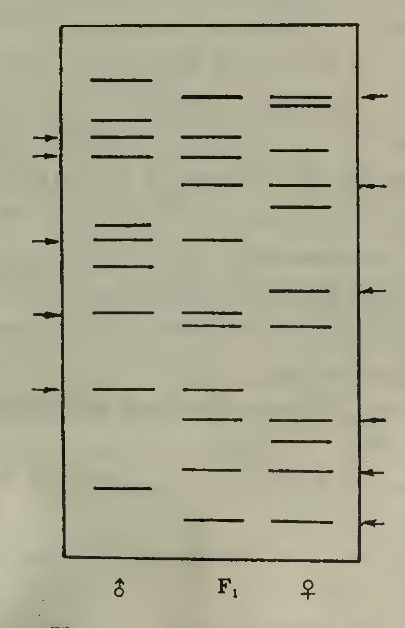


图 14-26　人的指纹分析示意图

♂为父亲的 DNA 指纹图谱；♀为母亲的 DNA 指纹图谱；F_1 为子/女可能出现的 DNA 指纹图谱。图左、右两侧的箭头表示子女 DNA 指纹图中分别与其父、母的 DNA 指纹中相对应的分子杂交条带。

RAPD 是随机扩增多态性 DNA (random amplified polymorphic DNA）的英文缩写。RAPD 是 1990 由 J. G. K. Williams 等发明的，是一种新奇的遗传标记。其方法是用随机序列的 9～10 个核苷酸的引物，对基因组的 DNA 进行 PCR 扩增，再进行电泳分离和溴化乙锭染色，多态性的产生可以由 DNA 序列中与互补寡聚核苷酸引物中一个碱基的差别形成。

RAPD 只产生显性标记（即有或没有），所以不能用来鉴定杂合体，但对自交系和回交群体来说这是不成问题的，因为所有或大多数的基因位点都是同源的。RAPD 也不能单独地用来对没有基因标记的生物进行遗传图谱分析，因为它所产生的多态性是随机的。

然而 RAPD 图谱比 RFLP 有一定的优点，人工合成的随机引物可在实验空间进行交换，这要比克隆 DNA 片段作为探针要容易和迅速得多，可以免去 DNA 分子杂交的手续，对大的基因组中低拷贝数序列也同样灵敏。RAPD 要求对每个分离群体中的标记所进行的 PCR 扩增

和电泳分离有高度的重复性，所以操作一般采用自动化。因此，RAPD 有可能成为将来遗传图谱构建中的一种普遍方法。

参考文献

1. 刘培楠，吴国利(主编). 基础分子生物学. 北京：高等教育出版社，1983.
2. 弗雷费尔德·D. 分子生物学(上、下册). 蔡武城等译. 北京：科学出版社，1988.
3. 卢因·B. 基因. 吴经才等译. 北京：科学出版社，1989.
4. 郭仁，张和君，董德祥. 分子细胞生物学. 北京：北京医科大学、协和医科大学联合出版社，1990.
5. Freifelder, D. Molecular Biology. (2nd ed.). Johns and Bartlett Publisher, Boston: 1987.
6. Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., Weiner, A. M. Molecular Biology of The Gene (4th ed.). Benjamin/Cummings, Menlo Park: 1987.
7. Watson, J. et al. Molecular Biology of The Gene Ⅱ. Benjamin/Cummings, Menlo Park: 1987.
8. Lewin, Ⅰ. Ⅴ. Genes. Ⅳ Oxford University Press, London: 1990.
9. Wettatein, D. Ⅴ. and Chua, N-H(蔡南海). Plant Molecular Biology. Plenum Press, New York: 1987.
10. Herrmann, R. G. and Larkins B A. Plant Molecular Biology 2 Plenum Press, New York: 1991.
11. Walker, J. M. & Gingold E. B. Molecular Biology & Biotechnology. (2nd ed.). Royal Society of Chemistry. Cambridge: 1992.
12. Voet, D. and Voet, J. G. Biochemistry. John Wiley & Sons, New York: 1990.
13. Mathews, C. K. and Van Holde., K. E. Biochemistry. Benjamin/Cummings, New York: 1990.
14. Stryer, L. Biochemistry. (2nd ed.). W. H. Freeman, San Francisco: 1987.
15. Fasman, G. D. Practical Handbook of Biochemistry and Molecular Biology. CRC Press, Inc. Boca Raton, Florida: 1989.
16. Burns, G. W. & Bottino, P J. The Science of Genetics. (6th ed.). Macmillan Publishing Co., New York: 1989.
17. Gardner, E. J., M. J. & Snustad, D. P. Principles of Genetics. (8th ed.). John Wiley & Sons, New York: 1991.
18. Russell, P. Genetics. (2th ed.). Scott, Foresman and Co., Boston: 1990.
19. Darnell, J., Lodish, H. & Baltimore, D. Molecular Cell Biology. W. H. Freeman, New York: 1990.
20. Avers, C. J. Molecular Cell Biology. Addison-Wesley, Reading, Massachusetts: 1986.
21. Tinoco, I. Jr., Sauer K., Wang., J. C. Physical Chemistry: Principles and Applications. Prentice-Hall, Inc., New Jersey: 1989.
22. Lin, E. C. C., Goldstein., R. & Syvanen, M. Bacteria, Plasmids and Phages an Introduction to Molecular Biology. Harvard University Press, Cambridge: 1984.
23. Brock, T. D., & Madigan, M. T. Biology of Micoroorganisms. (6th ed.). Prentice Hall, New Jersey: 1991.
24. Moat, A. G. & Foster, J. W. Microbial Physiology. (2th ed.). John Wiley, New York: 1988.
25. Adams, R. L. P., Knowler, J. T., & Leader, D. P. The Biochemistry of the Nucleic Acids. (11th ed.). Chapman & Hall, London: 1992.
26. Saenger, W. Principles of Nucleic Acid Structure. Springer-Verlag, New York: 1984.
27. Becker, Y. Replicution of Viral and Cellular Genomes Martinus Nijhoff Publishing, Boston: 1983.
28. Kornberg, A. DNA Replication. W. H. Freeman, San Francisco: 1983.
29. kornberg, A. Supplement to DNA Replication. W. H. Freeman, San Francisco: 1982.

30. Winnacker, E. L. From Genes to Clones VCH Verlag, Weinhein: 1987.
31. Collins, A., Downes, C. S. & Johnson. DNA Repair and Its Inhibition. IRL Press, Oxford: 1984.
32. Hames, B. D. & Higgins, S J. Transcription and Translation-A Practical Approach. IRL Press, Oxford: 1984.
33. Arnstein, H. R. V., & Cox, R. A. Protein Biosynthesis. IRL Press, Oxford University Press, 1991.
34. Tuite, M. F., Picard, M. & Bolotin-Fukuhara, M. Genetics of Translation. Springer-Verlag, Berlin. 1988.
35. Wright, T. R. F. Genetic Regulatory Hierarchies in Development. Academic Press, New York: 1990.
36. Goldberg, R. The Molecular Basis of Plant Development. Alan, R. Liss, Inc., New York: 1989.
37. Murphy, T. M. & Thompson, W. F. Molecular Plant Development. Prenetice Hall, New Jersey: 1988.
38. Heap, R. B., Prosser, C. G., & Lamming, G. E. Biotechnology in Growth Regulation. Butlerworths, London: 1989.
39. Dale, J. W. Molecular Genetics of Bacteria. John Wiley, New York: 1989.
40. Smith, K. M. & Ritchie, D. A. Introduction to Verology. Chapman and Hall, London: 1980.
41. Fraenkel-Conrat, H. & Kimball, P. C. Virology. Practice-Hall, New Jersey: 1982.
42. Strathern, J. N., Jones, E. W., & Broach, J. R. The Molecular Biology of the Yeast Saccharomycetes. Cold Spring Hanbor, New York: 1981.
43. Fincham, J. R. S. & Day P. R. Fungel Genetics. Philadelphia, Davis: 1971.
44. Adolph, K. W. Molecular Biology of Chromosome Function. Springer-Verlag, New York: 1989.
45. Lloyd, J. R. Genes and Chromosomes. Macmillan, London: 1986.
46. Brown, T. A. Essential Molecular Biology: A Practical Approach Ⅰ, Ⅱ. IRL Press, Oxford: 1991.
47. Glover, D. M. DNA Cloning: A Practial Approach Ⅰ, Ⅱ, Ⅲ. IRL Press, Oxford: 1985～1987.
48. Keller, G. H. & Manak., M. M. DNA Probes. Stockton Press, New York: 1989.
49. Maniatis, T., Frisch, E. F. & Sambrook. J. Molecular Cloning a Laboratory Manual Cold Spring Harbor Laboratory, New York: 1990.
50. 李德宝，徐平. 重组 DNA 的原理和方法. 杭州：浙江科学技术出版社，1993.
51. 沈珝琲，方福德主编. 真核基因表达调控. 北京：高等教育出版社，1996.

（王荣臣　龚小松）

索　引

E

F

G

H

I

J

K

L

M

N

P

Q

R

S

T

W

X

Y

Z



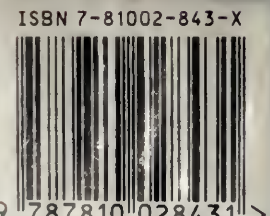


封面说明： DNA 三种形式的三维结构(左:B-DNA，中:TA-DNA，右:A-DNA。TA-DNA 为含有 TATA 框并能与 TATA 结合蛋白相结合的 DNA 片段。碱基为蓝色，脱氧核糖链骨架的氧及碳为红色，磷原子为黄色。



中华农业科教
基金资助图书

责任编辑:赵玉琴
封面设计:郑　川


ISBN 7-81002-843-X
9 787810 028431 >

定价:42.00 元